```
          GenCore version 5.1.6
   Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       July 21, 2004, 12:44:33 ; Search time 246 Seconds
                                    (without alignments)
                                    7697.124 Million cell updates/sec

Title:            US-09-270-437D-6
Perfect score:    3412
Sequence:         1 ggcagcggaggaggcgagga.........aacctgaaatgttattt 3412

Scoring table:    OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0

Searched:         682709 seqs, 277475446 residues

Word size :       0

Total number of hits satisfying chosen parameters:        1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        Issued Patents_NA:*
                  1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
                  2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
                  3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
                  4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
                  5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
                  6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                               SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3412	100.0	3412	3	US-09-061-709-6	Sequence 6, Appli
2	3412	100.0	3412	4	US-09-899-651-6	Sequence 6, Appli
3	2143	62.8	3283	3	US-09-061-709-8	Sequence 8, Appli
4	2143	62.8	3283	4	US-09-899-651-8	Sequence 8, Appli
C 5	34	1.0	317	4	US-09-621-976-13846	Sequence 13846, A
C 6	25	0.7	1149	4	US-09-644-460-36	Sequence 36, Appli
7	22	0.6	2224	3	US-09-261-885-1	Sequence 1, Appli
8	22	0.6	11049	4	US-10-204-708-23	Sequence 23, Appl
9	21	0.6	747	4	US-09-134-001C-2248	Sequence 2248, Ap
C 10	21	0.6	1602	1	US-08-075-533-1	Sequence 1, Appli
C 11	21	0.6	1602	2	US-08-948-176-1	Sequence 1, Appli
C 12	21	0.6	1602	5	PCT-US91-09160-1	Sequence 1, Appli
C 13	21	0.6	8961	4	US-10-204-708-80	Sequence 80, Appl
14	20	0.6	591	4	US-09-134-000C-2872	Sequence 2872, Ap
C 15	20	0.6	592	4	US-08-585-533A-27	Sequence 27, Appl
16	20	0.6	1539	4	US-09-364-230-29	Sequence 29, Appl
17	20	0.6	2199	4	US-09-708-725A-3	Sequence 3, Appli
C 18	20	0.6	2831	4	US-09-477-133A-117	Sequence 117, App
19	20	0.6	5428	4	US-09-625-972-21	Sequence 21, Appl
20	20	0.6	6070	4	US-10-204-708-10	Sequence 10, Appl
21	20	0.6	7174	4	US-08-961-527-189	Sequence 189, App
22	20	0.6	9641	4	US-09-625-972-22	Sequence 22, Appl
23	20	0.6	51259	3	US-08-781-891-209	Sequence 209, App
C 24	20	0.6	51259	4	US-09-618-166-209	Sequence 209, App
25	20	0.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
26	20	0.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 27	19	0.6	51	4	US-09-443-199C-203	Sequence 203, App

C 28	19	28	0.6	51	4	US-09-443-199C-204	Sequence 204, App
C 29	19	29	0.6	213	4	US-09-543-681A-1402	Sequence 1402, Ap
30	19	30	0.6	401	4	US-09-621-976-11383	Sequence 11383, A
31	19	31	0.6	1001	4	US-09-671-317-270	Sequence 270, App
32	19	32	0.6	1001	3	US-09-671-317-461	Sequence 461, App
33	19	33	0.6	1002	3	US-08-943-731-97	Sequence 97, Appl
34	19	34	0.6	1149	4	US-09-328-352-3312	Sequence 3312, Ap
35	19	35	0.6	1281	4	US-09-328-352-2290	Sequence 2290, Ap
C 36	19	36	0.6	1344	4	US-09-620-312D-937	Sequence 937, App
37	19	37	0.6	1443	4	US-09-802-927-2	Sequence 2, Appli
38	19	38	0.6	2186	2	US-08-878-546-9	Sequence 9, Appli
39	19	39	0.6	2238	1	US-08-173-436A-4	Sequence 4, Appli
C 40	19	40	0.6	2416	4	US-09-201-936-41	Sequence 41, Appl
C 41	19	41	0.6	2416	4	US-09-011-356-41	Sequence 41, Appl
C 42	19	42	0.6	2416	4	US-09-672-717-228	Sequence 228, App
43	19	43	0.6	2422	1	US-08-475-845-1	Sequence 1, Appli
C 44	19	44	0.6	2422	1	US-08-327-690-1	Sequence 1, Appli
C 45	19	45	0.6	2422	2	US-08-660-289-1	Sequence 1, Appli

```
                               ALIGNMENTS

RESULT 1
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
SEQ ID NO 6
;  LENGTH: 3412
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-061-709-6

Query Match             100.0%;  Score 3412;  DB 3;  Length 3412;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3412;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCGGGGAGCCGGGGCGCCTCTCGGGG  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCGGGGAGCCGGGGCGCCTCTCGGGG  60

Qy         61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGGGACCCCCGGTCACCGCGAC  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGGGACCCCCGGTCACCGCGAC  120

Qy        121 GACCTTCCGGCAGCTCTTTGGGGACAGGAGAGCTGCCCCTGGGGGACCAGGTCCTGCTGAAG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GACCTTCCGGCAGCTCTTTGGGGACAGGAGAGCTGCCCCTGGGGGACCAGGTCCTGCTGAAG  180

Qy        181 TCCGGCTACGGCCTTCGTGTGGACTACCCCGACCAGAACTGGGCCATCGAGACTCGAGACC  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCGGCTACGGCCTTCGTGTGGACTACCCCGACCAGAACTGGGCCATCGAGACTCGAGACC  240

Qy        241 CTCTCCGGGTAAAGTGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCGGGTAAAGTGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300

Qy        301 AAGTCAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG  360
```

```
Db   301 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         AAGCTAAGGAGGAGCAGGAGAAAATTCAGATTCGAAACATCCCTCCCTCACCTGCAGTGAGGGTG  360

Qy   361 TTGGATGGACTTTGGCTCAATATGGACAGTGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTGGCTCAATATGGACAGTGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC  420

Qy   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGAGCAAAAAATAGCCATG  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGAGCAAAAAATAGCCATG  480

Qy   481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT  540

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCTGGGGACCACTCTTCCCGG  600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCTGGGGACCACTCTTCCCGG  600

Qy   601 GAGCAAGGCCACGCCCCCTGGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCGG  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCCCTGGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCGG  660

Qy   661 ATCCTGGTCCCCACCCAGTTGTTGTGCCATCATCGGAAAGAAGGGGCTTGACCATAAAG  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTGTTGTGCCATCATCGGAAAGAAGGGGCTTGACCATAAAG  720

Qy   721 AACATCACTAAGCAGACCAGTCCCGGGGTAGTATATCCATAGAAAAGATGAGACTCTGGAGCT  780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCAGTCCCGGGGTAGTATATCCATAGAAAAGATGAGACTCTGGAGCT  780

Qy   781 GCAGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGACTTCTGAAGCATGCGCCATG  840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGACTTCTGAAGCATGCGCCATG  840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG  960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG  960

Qy   961 AAGAAAATTGAACATGAAAACAGGGACCAAGATAACAATCTCTTTGCAGGATTTGAGC  1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAAACAGGGACCAAGATAACAATCTCTTTGCAGGATTTGAGC  1020

Qy  1021 ATATACAACCCGGAAAGAACCATCACTGTGTGGAGTCACAGTTGAGGCCTGTGCCAGTGCT  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAACCATCACTGTGTGGAGTCACAGTTGAGGCCTGTGCCAGTGCT  1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGGCTGTTAAC  1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGGCTGTTAAC  1140

Qy  1141 CAACAAGCCAATCGAATCCCGGATACCTTGACATCCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CCCTTCACTACCCGGATACCTTGACATCCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200

Qy  1201 CTGTCCGTGTATCTCCACCACGAGGAGCAGCGGCCCCCGGCCTCCCCGCTGCCTACCAC  1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGTATCTCCACCACGAGGAGCAGCGGCCCCCGGCCTCCCCGCTGCCTACCAC  1260

Qy  1261 CCCTTCACTACCCACTCCGGATACTTCTCACCAGACGAGAGAGATTGAATCTCTTCATCCAGTTTGGCCCG  1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACCCACTCCGGATACTTCTCACCAGACGAGAGAGATTGAATCTCTTCATCCAGTTTGGCCCG  1320

Qy  1321 TTCCCCGCATCATCATCCCGTATCTCCACCACTCTTATCCAGAGCAGAGATTGTGAATTGTGCAGCTTTTTCATCCCAACCCAG  1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCCGCATCATCATCCCGTATCTCCACCACTCTTATCCAGAGCAGAGATTGTGAATTGTGCAGCTTTTTCATCCCAACCCAG  1380

Qy  1381 GCTGTGGGCGCCATCATCGGAAGAAGAGGGGGCACACATCAAACAGCTGGCGAGATTCGCC  1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCTGTGGGCGCCATCATCGGAAGAAGAGGGGGCACACATCAAACAGCTGGCGAGATTCGCC  1440

Qy  1441 GGAGCCTCTATCAAGATTGCCCCTCGCGAAGGCCCAGAGGCCGAAGGCTCAGCGAGTCAGCGGCCGAAGGATGGTCATC  1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTCTATCAAGATTGCCCCTCGCGAAGGCCCAGAGGCCGAAGGCTCAGCGAGTCAGCGGCCGAAGGATGGTCATC  1500

Qy  1501 ATCACCGGCCACCCGGAAGCCAGTTCAAGGTCAAGGCCCAGGACGGATCTTTGGAAACTGAAA  1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACCGGCCACCCGGAAGCCAGTTCAAGGTCAAGGCCCAGGACGGATCTTTGGAAACTGAAA  1560

Qy  1561 GAGGAAACTTCTTTAACCCAAAGAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC  1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAGGAAACTTCTTTAACCCAAAGAAGAAGAAGTCGAAGCGCATATCAGAGTGCCC  1620

Qy  1621 TCTTCCACAGCTGGCCCGGGTGATTGGCAAAGGTGCAAAGGTGGCAAGACCGTGACGAACTGCAGAAC  1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGCTGGCCCGGGTGATTGGCAAAGGTGCAAAGGTGGCAAGACCGTGACGAACTGCAGAAC  1680

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG  1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG  1740

Qy  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTCGACAGCGCAAGATCAGGGAA  1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTCGACAGCGCAAGATCAGGGAA  1800

Qy  1801 ATTGTACAACAGTGAAGCAGCAGGAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATTGTACAACAGTGAAGCAGCAGGAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860

Qy  1861 AGCAAGTGAAGGCTCCCACAGCAGCAGCACCACAAACAACGATGAATGTAGCCCTTCCAACAC  1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAAGGCTCCCACAGCAGCAGCACCACAAACAACGATGAATGTAGCCCTTCCAACAC  1920

Qy  1921 CTGACAGAATGAGACCAAAACGCCAGTCAGCGCAGCGCAGTGGGAGCAAACCAAAGACCCATCGAG  1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACCAAAACGCCAGTCAGCGCAGCGCAGTGGGAGCAAACCAAAGACCCATCGAG  1980

Qy  1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGAGACTCGCCGAGGCCCAGGGGGC  2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGAGACTCGCCGAGGCCCAGGGGGC  2040

Qy  2041 CGAGGAGGGGCCGGGGAGGTCAGCCGAGTTGCCAGAACCACCGCCTCCCGCC  2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGGCCGGGGAGGTCAGCCGAGTTGCCAGAACCACCGCCTCCCGCC  2100

Qy  2101 CCCCAGGGCTTCTGCAGCTTCAGCCATCCACTCCATCCTCCTGAA  2160
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGCTTCAGCCATCCACTCCATCCTCCTGAA  2160

Qy  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAAGCGTTCAAAGCCAAGC  2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAAGCGTTCAAAGCCAAGC  2220

Qy  2221 AAAATGCACACCCTTTTCGTGGCAAATCGTCTCGTGTACATGTGTGTACATAGAAA  2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTTCGTGGCAAATCGTCTCGTGTACATGTGTACATATTAGAAA  2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTCGGGTGCCTCGCAGCAGGGTTACACAGGGTAATATAT  2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTGCCTCGCAGCAGGGTTACACAGGGTAATATAT  2340

Qy  2341 TTTAGAAATAATATATCAAATAATTTTTTAATCAATTATTAATTTT  2400
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATATCAAATAATTCAATTTTTAATCAATTATTAATTTT  2400

Qy  2401 TTTTTCTTTTTAAAGAATAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTTAAAGAATAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460

Qy  2461 GTCTCACGGTGTAGAGAGGAGGACTTTGAGCGCACCCCGCACCCCACAAAATTCACCCAGAGGGAAAT  2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGGAGGACTTTGAGCGCACCCCGCACCCCACAAAATTCACCCAGAGGGAAAT  2520
```

```
APPLICANT: Old, Lloyd J.
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
TITLE OF INVENTION: Associated
TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
FILE REFERENCE: LUD 5538
CURRENT APPLICATION NUMBER: US/09/899,651
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US/09/061,709
PRIOR FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 8
SEQ ID NO 6
  LENGTH: 3412
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
US-09-899-651-6

Query Match            100.0%;   Score 3412;   DB 4;   Length 3412;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 3412;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy      1  GGCAGCGGAGGAGGCGGAGGAGCGCCGGGGTACCGGCCGGGGAGCCGGCTCTCGGGGG  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GGCAGCGGAGGAGGCGGAGGAGCGCCGGGGTACCGGCCGGGGAGCCGGCTCTCGGGGG  60

Qy     61  AAGAGACGATGATGAACAAGCTTTACATCGGGACTTTACATCGGGACTGAGCCCCGCCGAC  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  AAGAGACGATGATGAACAAGCTTTACATCGGGACTTTACATCGGGACTGAGCCCCGCCGAC  120

Qy    121  GACTTCCGGGACGAGCTTCTTTGGGGACAGGAAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GACTTCCGGGACGAGCTTCTTTGGGGACAGGAAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG  180

Qy    181  TCCGGCTACCGCCTTCGTGGACTACCCGACCAGAACTGGGCCATCCGGCCATCCGAGACC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TCCGGCTACCGCCTTCGTGGACTACCCGACCAGAACTGGGCCATCCGGCCATCCGAGACC  240

Qy    241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCAGATTGATTACTCAGTCTTAA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCAGATTGATTACTCAGTCTTAA  300

Qy    301  AAGCTAAGGAGCAGAGAAATTCAGATTCGGACAGTGGAGAATGTGGACAAGTCAACACAGAC  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AAGCTAAGGAGCAGAGAAATTCAGATTCGGACAGTGGAGAATGTGGACAAGTCAACACAGAC  360

Qy    361  TTGGATGGACTTTTGGCTCAATATGGACAGTGAAAGCTGATAG  420
           ||||||||||||||||||||||||||||||||||||||||||||
Db    361  TTGGATGGACTTTTGGCTCAATATGGACAGTGAAAGCTGATAG  420

Qy    421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAGCAACAGAAATAGCCATG  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAGCAACAGAAATAGCCATG  480

Qy    481  GAGAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GAGAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT  540

Qy    541  GAAGAGGTGAGCTCCCCTTGCGCCCCCTGGGGACCACTCTTCCCGG  600
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GAAGAGGTGAGCTCCCCTTGCGCCCCCTGGGGACCACTCTTCCCGG  600

Qy    601  GAGCAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GAGCAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG  660

Qy    661  ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGCTTGACCATAAAG  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGCTTGACCATAAAG  720

Qy    721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT  780
```

```
Qy   2521  CTCGTCGGAGGACACTCACGGCAGTTCGGATCACCTGTGTATGTCAACAGAAGGGATA  2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521  CTCGTCGGAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA  2580

Qy   2581  CCGTCTCCTTGAAGAGGAAACTCTGCACTCCTCGTCCTCAGCTCATCACACCCATT  2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  CCGTCTCCTTGAAGAGGAAACTCTGCACTCCTCGTCCTCAGCTCATCACACCCATT  2640

Qy   2641  TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATATAATTCTCTGTCT  2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641  TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATATAATTCTCTGTCT  2700

Qy   2701  CTCTCTGTTTATCTCTCCCCCCTCCCCTTCTCTCTCTCCATCTCCATTCTTTGAA  2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701  CTCTCTGTTTATCTCTCCCCCCTCCCCTTCTCTCTCTCCATCTCCATTCTTTGAA  2760

Qy   2761  TTTCCTCATCCCTCCATCTCAATCCCGTATCCAGCACCCCCCCCCCCCCCAGGCAAAGCA  2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761  TTTCCTCATCCCTCCATCTCAATCCCGTATCCAGCACCCCCCCCCCCCCCAGGCAAAGCA  2820

Qy   2821  GTGCTCTGAGTATCACATCACACAAAAGGAACACAAAGCGAAACACAGCCTCAA  2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  GTGCTCTGAGTATCACATCACACAAAAGGAACACAAAGCGAAACACAGCCTCAA  2880

Qy   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG  2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG  2940

Qy   2941  AGGAAAACACAGAACCCACCAACAACCAATCAACCAAACAAAGAAAAAATTCCACAATG  3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  AGGAAAACACAGAACCCACCAACAACCAATCAACCAAACAAAGAAAAAATTCCACAATG  3000

Qy   3001  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060

Qy   3061  ATTCCTTTCTTTAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  ATTCCTTTCTTTAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120

Qy   3121  CAGGGGCCTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAGAGCTACCTCAG  3180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  CAGGGGCCTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAGAGCTACCTCAG  3180

Qy   3181  GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG  3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG  3240

Qy   3241  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT  3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT  3300

Qy   3301  GCCAGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTCTGAATTGTCTTCCGCT  3360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  GCCAGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTCTGAATTGTCTTCCGCT  3360

Qy   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
```

```
RESULT 2
US-09-899-651-6
; Sequence 6, Application US/09899651
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
```

```
Qy    781 GCAGAGAGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCGCCATG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGAGAGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCGCCATG 840

Qy    841 ATTCTTGAAATCATGCAGAAAGAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAGAAAGAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960

Qy    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTGGCTGTTAAC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTGGCTGTTAAC 1140

Qy   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy   1201 CTGTCCGTGCTATCTCCACCACTCCACCAGCAGGGCCCCGGAGCTCCCCGCTGCCCCCTACCAC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGCTATCTCCACCACTCCACCAGCAGGGCCCCGGAGCTCCCCGCTGCCCCCTACCAC 1260

Qy   1261 CCCTTCACTACCACCCGGATACTTCTCCAGCAGGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTACCACCCGGATACTTCTCCAGCAGGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1320

Qy   1321 TTCCCGCATCATCACTCTTATCCAGACAGGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCGCATCATCACTCTTATCCAGACAGGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380

Qy   1381 GCTGTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGTCTGGCGAGATTCGCC 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGTCTGGCGAGATTCGCC 1440

Qy   1441 GGAGCCTCTATCAAGATTGCCCTGCGGAAGCCCCAGACTTCAGCGAAAGGATGGTCATC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGCCTCTATCAAGATTGCCCTGCGGAAGCCCCAGACTTCAGCGAAAGGATGGTCATC 1500

Qy   1501 ATCACCGGGCACCGGAAGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGGCACCGGAAGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560

Qy   1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCTGGAAGCCCATATCAGAGTGCCC 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCTGGAAGCCCATATCAGAGTGCCC 1620

Qy   1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACTGCAGAAC 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACTGCAGAAC 1680

Qy   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAAGTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAAGTG 1740

Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACGCCGCAAGATCAGGGAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACGCCGCAAGATCAGGGAA 1800

Qy   1801 ATTGTACAACAGGTTGAAGCAGTGAAGCAGCAGCGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGGTTGAAGCAGTGAAGCAGCAGCGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860

Qy   1861 AGCAAGTGAGGCTCCCACAGGCACCACCAGCAGCCAGCAGCCAGATGAGGAAGCTCCAGAGACCATCTGAG 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCTCCCACAGGCACCACCAGCAGCCAGCAGCCAGATGAGGAAGCTCCAGAGACCATCTGAG 1920

Qy   1921 CTGACAGAATGAGACCAAACGCACCAGCAGCCAGATGAGGAGCAAACCAAACCATCTGAG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACAGAATGAGACCAAACGCACCAGCAGCCAGATGAGGAGCAAACCAAACCATCTGAG 1980

Qy   1981 GAATAGAGAGTCTGCGGAGGCGGAGGCGGCGGCCAGGGACTCTGCGCGAGGCCCTGAGGAGACCACCCAGGGGC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATAGAGAGTCTGCGGAGGCGGAGGCGGCGGCCAGGGACTCTGCGCGAGGCCCTGAGGAGACCACCCAGGGGC 2040

Qy   2041 CGAGGAGGGGCGGCGGGGAAGTCAGCCAGTTTGCCAGAACCACCAGGAGAACCACCAGCCGGGCTCCCCGCCC 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGAGGAGGGGCGGCGGGGAAGTCAGCCAGTTTGCCAGAACCACCAGGAGAACCACCAGCCGGGCTCCCCGCCC 2100

Qy   2101 CCCCAGGGCTTCTGCAGCTTCAGCGCCATCCAGTTCACCATCCACTTCACCATCCACTCGGATCTCTCCTGAA 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCCAGGGCTTCTGCAGCTTCAGCGCCATCCAGTTCACCATCCACTTCACCATCCACTCGGATCTCTCCTGAA 2160

Qy   2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220

Qy   2221 AAAATGCACACCCTTTTCGTGTGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AAAATGCACACCCTTTTCGTGTGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280

Qy   2281 GGGAAGATGTTAAGATAATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GGGAAGATGTTAAGATAATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATAT 2340

Qy   2341 TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT 2400

Qy   2401 TTTTTCTTTTAAAGAGAAAGCAGAGGCTTTTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTTTTCTTTTAAAGAGAAAGCAGAGGCTTTTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy   2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCGAGAGGGAAAT 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCGAGAGGGAAAT 2520

Qy   2521 CTCGTCCGAAGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGGAAGGGATA 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTCGTCCGAAGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGGAAGGGATA 2580

Qy   2581 CCGTTCTCCTTGGAAGAGAAACTCTGTCACTCGTCACTCCTCATGCCTGCTCAGCTCACTACACCCATT 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCGTTCTCCTTGGAAGAGAAACTCTGTCACTCGTCACTCCTCATGCCTGCTCAGCTCACTACACCCATT 2640

Qy   2641 TCTCTTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGCTATATAATTCTCTGTCT 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TCTCTTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGCTATATAATTCTCTGTCT 2700

Qy   2701 CTCTCGTTTATCTCTCCCCCTTCTTCTCCATCTCCATTCTTTGAA 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCTCGTTTATCTCTCCCCCTTCTTCTCCATCTCCATTCTTTGAA 2760

Qy   2761 TTTCCTCATCCCTCAATCCGTATCTCACGGCACCCCCCCGTATCCCGAAAGCAAAGCA 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TTTCCTCATCCCTCAATCCGTATCTCACGGCACCCCCCCGTATCCCGAAAGCAAAGCA 2820

Qy   2821 GTGCTCTGAGTATCACATCACAACAAAGGACAAAAGCGAAACACAACCAGCCTCAA 2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GTGCTCTGAGTATCACATCACAACAAAGGACAAAAGCGAAACACAACCAGCCTCAA 2880

Qy   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAGCGC 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAGCGC 2940

Qy   2941 AGGAAAACAGGAACCCACCAGACCAAACCAATCAACCAATCAACAAAGAAAAATTCCACAATG 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2941 AGGAAAACAGGAACCCACCAACCAATCAACCAAACAAGAAAAATTCCACAATG 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1321 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGGAAGGATGGTCATCATCACCGGG 1380

Db       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3001 AAAGAATGTATTTGTCTTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
Qy  1510 CCACCGGAAGCCCAGTTCAAGGCCCAGACGTTTGGGAAACTGAAAGAGAGGAAAAC 1569

Db  3001 AAAGAATGTATTTGTCTTTTTGCATTTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CCACCGGAAGCCCAGTTCAAGGCCCAGACGTTTGGGAAACTGAAAGAGAGGAAAAC 1440

Qy  3061 ATTCCTTTCTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
Qy  1570 TTCTTTAACCCCAAAGAAGAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACA 1629

Db  3061 ATTCCTTTCTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
Db  1441 TTCTTTAACCCCAAAGAAGAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACA 1500

Qy  3121 CAGGGGGTTAAATTCACAGATTTTTTAACGAGAAAACACAGAAGAGCTACCTCAG 3180
Qy  1630 GCTGCGCGGTGATTGGCAAAGGTGGCAAACCGTGAACGACGACTTAACCAGT 1689

Db  3121 CAGGGGGTTAAATTCACAGATTTTTTAACGAGAAAACACAGAAGAGCTACCTCAG 3180
Db  1501 GCTGCGCGGTGATTGGCAAAGGTGGCAAACCGTGAACGACGACTTAACCAGT 1560

Qy  3181 GTGTTTTTTTACCTCAGCACCTTGCTCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
Qy  1690 GCAGAAGTCATCGTGCCTGCGTCTGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1749

Db  3181 GTGTTTTTTTACCTCAGCACCTTGCTCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
Db  1561 GCAGAAGTCATCGTGCCTGCGTCTGCCAGATGAAAATGAGGAAGTGATCGTCAGA 1620

Qy  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
Qy  1750 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA 1809

Db  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
Db  1621 ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA 1680

Qy  3301 GCCAGCCTGGAGAAGTGACAGTGCAAGTCCAAGTGTCGCAACAGCAGCTGTCTTCCGCT 3360
Qy  1810 CAGGTGAAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCCCTCACAGCGCCAGCAAGTGA 1869

Db  3301 GCCAGCCTGGAGAAGTGACAGTGCAAGTCCAAGTGTCGCAACAGCAGCTGTCTTCCGCT 3360
Db  1681 CAGGTGAAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCCCTCACAGCGCCAGCAAGTGA 1740

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
Qy  1870 GGCTCCCACAGGCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCTGACACAGAA 1929

Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
Db  1741 GGCTCCCACAGGCACCAGCAAACAACGGATGAATGTAGCCCTTCCAACACCTGACACAGAA 1800

Qy  1930 TGAGACCAAACGCAGCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGAATGAGAA 1989

Db  1801 TGAGACCAAACGCAGCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGAATGAGAA 1860

Qy  1990 GTCTGCGGAGGCGGCCGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGCCGAGGAGGG 2049

Db  1861 GTCTGCGGAGGCGGCCGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGCCGAGGAGGG 1920

Qy  2050 GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCCCCAGGGGC 2109

Db  1921 GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCCCCAGGGGC 1980

Qy  2110 TTCTGCAGGTTCAGCATCCACTTCACCATCCACTCCTCCACGATCTCTCCGACTCCCACGA 2169

Db  1981 TTCTGCAGGTTCAGCATCCACTTCACCATCCACTCCTCCACGATCTCTCCGACTCCCACGA 2040

Qy  2170 CGCATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2229

Db  2041 CGCATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC 2100

Qy  2230 ACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTGTACATATATTAGAAAGGGAAGATG 2289

Db  2101 ACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTGTACATATATTAGAAAGGGAAGATG 2160

Qy  2290 TTAAGATATGTGGCCTGTGGGTACACAGGTGCCTGCGAGCGTGACAGGTTAATATTTTAGAAAT 2349

Db  2161 TTAAGATATGTGGCCTGTGGGTACACAGGTGCCTGCGAGCGTGACAGGTTAATATATTTTAGAAAT 2220

Qy  2350 AATATATCAAATAACTCAATCCAATTTTAATCAATTATTAATTTTTTTTCTTT 2409

Db  2221 AATATATCAAATAACTCAATCCAATTTTAATCAATTATTAATTTTTTTTCTTT 2280

Qy  2410 TTAAAGAAAGCAGCAGCTTTGAGGCTTTAAAGATAAAGTCTTTGGGAGGTCTCACGG 2469

Db  2281 TTAAAGAAAGCAGCAGCTTTGAGGCTTTAAAGATAAAGTCTTTGGGAGGTCTCACGG 2340

Qy  2470 TGTAGAGAGGAGCTTTGAGGCCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2529

Db  2341 TGTAGAGAGGAGCTTTGAGGCCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA 2400

Qy  2530 AGGACACTCACGGACAGTTCTGGATCACCTGTGTATGTCAACGAAGGGATACCGTCTCCT 2589
```

RESULT 3
US-09-061-709-8

; Sequence 8, Application US/09061709B
; Patent No. 5297364
; GENERAL INFORMATION:
;   APPLICANT: Chen, Yao-Tseng
;   APPLICANT: Gure, Ali
;   APPLICANT: Tsang, Solam
;   APPLICANT: Stockert, Elisabeth
;   APPLICANT: Jager, Elke
;   APPLICANT: Knuth, Alexander
;   APPLICANT: Old, Lloyd J.
;   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
;   TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;   FILE REFERENCE: LUD 5538
;   CURRENT APPLICATION NUMBER: US/09/061,709B
;   CURRENT FILING DATE: 1998-04-17
;   SEQ ID NOS: 8
;   NUMBER OF SEQ ID NOS: 8
;   LENGTH: 3283
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-061-709-8

Query Match   62.8%; Score 2143; DB 3; Length 3283;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  1270 ACCCACTCCGGATACTTCTCCAGCGCAGAGCAGGAGATTCGCCGGAGCCTGTGGGC 1329
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 ACCCACTCCGGATACTTCTCCAGCGCAGAGCAGGAGATTCGCCGGAGCCTGTGGGC 1200

Qy  1330 CATCACTCTATCCAGAGCAGGAGATTGTGAATCTTCTTCAGAACTTTGGGGAGGTCTCACGG 1389
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CATCACTCTATCCAGAGCAGGAGATTGTGAATCTTCTTCAGAACTTTGGGGAGGTCTGGGGC 1260

Qy  1390 GCCATCATCGGGAAGAGGGGGCCACACATCAAACAGCTGGGCGAGATTCGCCGGAGCCTCT 1449
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GCCATCATCGGGAAGAGGGGGCCACACATCAAACAGCTGGGCGAGATTCGCCGGAGCCTCT 1320

Qy  1450 ATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGGAAGGATGGTCATCATCACCGGG 1509
```

```
;  TITLE OF INVENTION:  Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
;  SEQ ID NO 8
;  LENGTH: 3283
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-899-651-8

Query Match             62.8%;  Score 2143;  DB 4;  Length 3283;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2143;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy       1270 ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGCCCGTTCCCGCAT 1329
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGCCCGTTCCCGCAT 1200

Qy       1330 CATCACTCCTATCCAGACAGCAGGAGATTGTGAATCTCTCATCCCAACCAGGCTGTGGGC 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CATCACTCCTATCCAGACAGCAGGAGATTGTGAATCTCTCATCCCAACCAGGCTGTGGGC 1260

Qy       1390 GCCATCATCGGAAGAAGGAGGGGGCCACACATCAAACAGCTGCCGGAGATTCGCCGGAGCCTCT 1449
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCCATCATCGGAAGAAGGAGGGGGCCACACATCAAACAGCTGCCGGAGATTCGCCGGAGCCTCT 1320

Qy       1450 ATCAAGATTGCCCCTGCGGAAGCCCCAGACGTCAGCGGAGGCCCAGAGATGGTCATCATCACCGGG 1509
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATCAAGATTGCCCCTGCGGAAGCCCCAGACGTCAGCGGAGGCCCAGAGATGGTCATCATCACCGGG 1380

Qy       1510 CCACCGGAGCCCAGTTCAAGGCCAGGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCACCGGAGCCCAGTTCAAGGCCAGGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1440

Qy       1570 TTCTTTAACCCAAAGAAGAAGAGTGAAGCGCATATCAGAGTGCCCTTCTTCCACA 1629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTCTTTAACCCAAAGAAGAAGAGTGAAGCGCATATCAGAGTGCCCTTCTTCCACA 1500

Qy       1630 GCTGGCCCGGTGATTGGCAAGGTGGCAAGACCGTGCAAGAACTGCAGAACTTAACCAGT 1689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GCTGGCCCGGTGATTGGCAAGGTGGCAAGACCGTGCAAGAACTGCAGAACTTAACCAGT 1560

Qy       1690 GCAGAAGTCATCGTGCCTCGTCGTGACCAAAACGCCAGATGAGGAGTCGATCGTCAGA 1749
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCAGAAGTCATCGTGCCTCGTCGTGACCAAAACGCCAGATGAGGAGTCGATCGTCAGA 1620

Qy       1750 ATTATCGGGCACTTCTTTGCTCAACAGCTGCTGAATTGTCTTCCCGCTAGCCAAGAA 1809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTATCGGGCACTTCTTTGCTCAACAGCTGCTGAATTGTCTTCCCGCTAGCCAAGAA 1680

Qy       1810 CAGGTGAAGCAGCAGGAGCAGCAAAACAACACGGATGAATGTAGCCCTTCAACACCTGACAGAA 1869
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAGGTGAAGCAGCAGGAGCAGCAAAACAACACGGATGAATGTAGCCCTTCAACACCTGACAGAA 1740

Qy       1870 GGCTCCCACAGGCACCACGGCACAGCAAAACACAACAACACGGAGAATGAAGA 1929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GGCTCCCACAGGCACCACGGCACAGCAAAACACAACAACACGGAGAATGAAGA 1800

Qy       1930 TGAGACCAAACGCAGCCAGCCAGATCGGAGCCAAACCAAAGACCATCTGAGGAATGAGAA 1989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGAGACCAAACGCAGCCAGCCAGATCGGAGCCAAACCAAAGACCATCTGAGGAATGAGAA 1860

Qy       1990 GTCTGCGGAGGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGCCGAGGAGGG 2049
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCTGCGGAGGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCGAGGGGCCGAGGAGGG 1920

Qy       2050 GCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCGCCCCCCAGGGGC 2109
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCGCCCCCCAGGGGC 1980

Db       2401 AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGAGGATACCGTCTCCT 2460

Qy       2590 TGAAGAGGAAACTCTGTCACTCCTCAGCTCATCACACCCATTTCTCTTTGC 2649
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGAAGAGGAAACTCTGTCACTCCTCAGCTCATCACACCCATTTCTCTTTGC 2520

Qy       2650 TTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT 2709
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT 2580

Qy       2710 TATCTCTCCCCCCCCCCCCCTTCTTCTCCATCTCCATCTCCATTCTTTGAATTTCCTCAT 2769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TATCTCTCCCCCCCCCCCCTTCTTCTCCATCTCCATCTCCATTCTTTGAATTTCCTCAT 2640

Qy       2770 CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2829
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2700

Qy       2830 GTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAACTTACACTT 2889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAACTTACACTT 2760

Qy       2890 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAGAGGAAAACA 2949
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAGAGGAAAACA 2820

Qy       2950 GGAACCCACCAAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAAATGT 3009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 GGAACCCACCAAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAAATGT 2880

Qy       3010 ATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 ATTTTGTCTTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 2940

Qy       3070 TTTAAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCCGTT 3129
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TTTAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCCGTT 3000

Qy       3130 AAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTA 3189
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 AAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTA 3060

Qy       3190 CCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTGTTGGAGCAT 3249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 CCTCAGCACCTTGCTCTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTGTTGGAGCAT 3120

Qy       3250 TTTTTATTTTTTTAATAAAAAATGAGTTGGAAAAAATTAAGATTATCAACTGCCAGCCTG 3309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 TTTTTATTTTTTTAATAAAAAATGAGTTGGAAAAAATTAAGATTATCAACTGCCAGCCTG 3180

Qy       3310 GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCCGCTAGCCAAGAA 3369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCCGCTAGCCAAGAA 3240

Qy       3370 CCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

RESULT 4
US-09-899-651-8

; Sequence 8, Application US/09899651
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated

```
Qy   2110  TTCTGCAGGCTTCAGCCATCCACTTCACCATCACCTCGGATCTCCTGAACTCCCACGA  2169
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  TTCTGCAGGCTTCAGCCATCCACCATCCACCATCACCTCGGATCTCCTGAACTCCCACGA  2040

Qy   2170  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2229
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2100

Qy   2230  ACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2289
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  ACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2160

Qy   2290  TTAAGATATGTGGCCTGTGGGTTACCACAGGCGTGCCTGCAGCGGTAATATATTTAGAAAT  2349
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  TTAAGATATGTGGCCTGTGGGTTACCACAGGCGTGCCTGCAGCGGTATATATTTTAGAAAT  2220

Qy   2350  AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTAATTTTTTTTCTTT  2409
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTAATTTTTTTTTCTTT  2280

Qy   2410  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG  2469
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCACGG  2340

Qy   2470  TGTAGAGGAGGAGCTTTGAGGCACCAGGCCACCCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2529
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  TGTAGAGGAGGAGCTTTGAGGCACCAGGCCACCCCGCACAAAATTCACCCAGAGGGAAATCTCGCGA  2400

Qy   2530  AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGGGATACCGTCTCCT  2589
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  AGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAAGGGGATACCGTCTCCT  2460

Qy   2590  TGAAGAGGAAACTCTGTCACTCCCATGCCCTGTCAGCTCATACACCCATTTCTCTTTGC  2649
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461  TGAAGAGGAAACTCTGTCACTCCCATGCCCTGTCAGCTCATACACCCATTTCTCTTTGC  2520

Qy   2650  TTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCCTGTCTGTCTCTGTT  2709
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521  TTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCCTGTCTGTCTCTGTT  2580

Qy   2710  TATCTCTCCCCTCCCCCTCCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCTCAT  2769
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  TATCTCTCCCCTCCCCCTCCCCCTCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCTCAT  2640

Qy   2770  CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCAGGCAAAGCCAGTGCTCTGA  2829
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641  CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCAGGCAAAGCCAGTGCTCTGA  2700

Qy   2830  GTATCACATCACACAAAAGGAACAAAAGGGAAAACACACAACCAGCCTCAACTTACACTT  2889
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701  GTATCACATCACACAAAAGGAACAAAAGGGAAAACACAACCAGCCTCAACTTACACTT  2760

Qy   2890  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACA  2949
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACA  2820

Qy   2950  GGAACCCACCAAACCAACCAATCAACCAACAAAGAAAAAATTCCACAATGAAAGAATGT  3009
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  GGAACCCACCAAACCAACCAATCAACCAACAAAGAAAAAATTCCACAATGAAAGAATGT  2880

Qy   3010  ATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC  3069
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  ATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC  2940

Qy   3070  TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCCAGGGCGTT  3129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCCAGGGCGTT  3000

Qy   3130  AAATTCACAGATTTTTTTAACGAGAAAAACACACAGAAGAGAGCTACCTCAGGTGTTTTTA  3189
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  AAATTCACAGATTTTTTTTAACGAGAAAAACACACCAGAAGAGAGCTACCTCAGGTGTTTTTTA  3060
```

```
Qy   3190  CCTCAGCACCTTGCTCTGTCGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT  3249
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  CCTCAGCACCTTGCTCTGTCGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCAT  3120

Qy   3250  TTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTG  3309
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  TTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTG  3180

Qy   3310  GAGAAGGTGACAGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAA  3369
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  GAGAAGGTGACAGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAA  3240

Qy   3370  CCNATATGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT  3412
           ||||||||||||||||||||||||||||||||||||||||
Db   3241  CCNATATGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT  3283
```

RESULT 5
US-09-621-976-13846/c
; Sequence 13846, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 13846
;   LENGTH: 317
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-13846

  Query Match   1.0%; Score 34; DB 4; Length 317;
  Best Local Similarity   100.0%; Pred. No. 3.3e-06;
  Matches   34; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2595 AGGAAACTCTGTCACTCCTCCATGCCGTGTCTAGCT 2628
           ||||||||||||||||||||||||||||||||||||
Db   216 AGGAAACTCTGTCACTCCTCCATGCCGTGTCTAGCT 183

RESULT 6
US-09-644-460-36
; Sequence 36, Application US/09644460
; Patent No. 6657053
; GENERAL INFORMATION:
;   APPLICANT: Fisher, Paul B.
;   TITLE OF INVENTION: Reciprocal Subtraction Differential
;   TITLE OF INVENTION: Display
;   FILE REFERENCE: 34587-C-PCT-USA
;   CURRENT APPLICATION NUMBER: US/09/644,460
;   CURRENT FILING DATE: 2000-08-23
;   PRIOR APPLICATION NUMBER: PCT/US99/04323
;   PRIOR FILING DATE: 1999-02-26
;   PRIOR APPLICATION NUMBER: US 09/197,889
;   PRIOR FILING DATE: 1998-11-23
;   PRIOR APPLICATION NUMBER: US 09/185,115
;   PRIOR FILING DATE: 1998-11-03
;   PRIOR APPLICATION NUMBER: US 09/032,684
;   PRIOR FILING DATE: 1998-02-27
;   NUMBER OF SEQ ID NOS: 42
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-644-460-36

Query Match   0.7%;   Score 25;   DB 4;   Length 1149;
Best Local Similarity   100.0%;   Pred. No. 0.066;
Matches   25;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    2957 ACCAAACCACCAACCAATCAACCAAACA 2981
                |||||||||||||||||||||||||||
Db     932 ACCAAACCACCAACCAATCAACCAAACA 956
```

RESULT 7
US-09-261-855-1
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
;   APPLICANT: Ross, Jeffrey
;   TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;   TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;   FILE REFERENCE: 960296.95131
;   CURRENT APPLICATION NUMBER: US/09/261,855A
;   CURRENT FILING DATE: 1999-03-03
;   NUMBER OF SEQ ID NOS: 46
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2224
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-261-855-1

Query Match   0.6%;   Score 22;   DB 3;   Length 2224;
Best Local Similarity   100.0%;   Pred. No. 1.8;
Matches   22;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    1376 CCCAGGCTGTGGGCGGCCATCAT 1397
                ||||||||||||||||||||||
Db    1371 CCCAGGCTGTGGGCGGCCATCAT 1392
```

RESULT 8
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;   TITLE OF INVENTION: by Assessing DNA Methylation
;   FILE REFERENCE: 5013 1012
;   CURRENT APPLICATION NUMBER: US/10/204,708
;   CURRENT FILING DATE: 2003-05-06
;   PRIOR APPLICATION NUMBER: PCT/EP01/03971
;   PRIOR FILING DATE: 2001-04-06
;   PRIOR APPLICATION NUMBER: DE 10019058.8
;   PRIOR FILING DATE: 2000-04-06
;   PRIOR APPLICATION NUMBER: DE 10019173.8
;   PRIOR FILING DATE: 2000-04-07
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

Query Match   0.6%;   Score 22;   DB 4;   Length 11049;
Best Local Similarity   100.0%;   Pred. No. 1.8;
Matches   22;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    3248 ATTTTTTATTTTTTAATAAA 3269
                |||||||||||||||||||
Db     680 ATTTTTTATTTTTTAATAAA 701
```

RESULT 9
US-09-134-001C-2248/c
; Sequence 2248, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;   APPLICANT: Lynn Doucette-Stamm et al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;   TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC-007
;   CURRENT APPLICATION NUMBER: US/09/134,001C
;   CURRENT FILING DATE: 1998-08-13
;   PRIOR APPLICATION NUMBER: US 60/064,964
;   PRIOR FILING DATE: 1997-11-08
;   PRIOR APPLICATION NUMBER: US 60/055,779
;   PRIOR FILING DATE: 1997-08-14
;   NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2248
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2248

Query Match   0.6%;   Score 21;   DB 4;   Length 747;
Best Local Similarity   100.0%;   Pred. No. 5.3;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    3248 ATTTTTTATTTTTTAATAA 3268
                ||||||||||||||||||
Db     175 ATTTTTTATTTTTTAATAA 155
```

RESULT 10
US-08-075-533-1/c
; Sequence 1, Application US/08075533
; Patent No. 5530186
; GENERAL INFORMATION:
;   APPLICANT: Hitz, William D.
;   APPLICANT: Yadav, Narendra S.
;   TITLE OF INVENTION: Nucleotide Sequences of Soybean Acyl-ACP
;   TITLE OF INVENTION: Thioesterase Genes
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. du Pont de Nemours and Company
;   STREET: 1007 Market Street
;   CITY: Wilmington
;   STATE: Delaware
;   COUNTRY: U.S.A
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/075,533
;   FILING DATE:
;   CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/631,264
;   FILING DATE: 20-DEC-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Morrissey, Bruce W.
;   REGISTRATION NUMBER: 30,663
;   REFERENCE/DOCKET NUMBER: CR-8926-A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (302) 992-4927
;   TELEFAX: (302) 892-7949
;   TELEX: 835420

INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1602 base pairs
    TYPE: Nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    ORGANISM: Glycine max
    STRAIN: Cultivar Wye
    DEVELOPMENTAL STAGE: Early seed fill
    HAPLOTYPE: Diploid
    TISSUE TYPE: Cotyledon
    ORGANELLE: Nucleus
  IMMEDIATE SOURCE:
    LIBRARY: cDNA to mRNA
    CLONE: 22B
  FEATURE:
    NAME/KEY: mat_peptide
    LOCATION: 271..1206
    IDENTIFICATION METHOD: Catalytically active when
    IDENTIFICATION METHOD: expressed in E. coli
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 106..1209
US-08-075-533-1

Query Match   0.6%; Score 21; DB 1; Length 1602;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3251 TTTTTATTTTTTAATAAAAA 3271
          |||||||||||||||||||
Db   87 TTTTTATTTTTTAATAAAAA 67

RESULT 11
US-08-948-176-1/c
; Sequence 1, Application US/08948176
; Patent No. 5945585
; GENERAL INFORMATION:
  APPLICANT: HITZ, WILLIAM D.
  APPLICANT: YADAV, NARENDRA S.
  TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES
  TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT
  TITLE OF INVENTION: OIL COMPOSITION
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
    STREET: 1007 MARKET STREET
    CITY: WILMINGTON
    STATE: DELAWARE
    COUNTRY: UNITED STATES OF AMERICA
    ZIP: 19898
  COMPUTER READABLE FORM:
    MEDIUM TYPE: DISKETTE, 3.50 INCH
    COMPUTER: IBM PC COMPATIBLE
    OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
    SOFTWARE: MICROSOFT WORD VERSION 7.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/948,176
    FILING DATE:
    CLASSIFICATION: 800
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/631,264
    FILING DATE: DECEMBER 20, 1990
  ATTORNEY/AGENT INFORMATION:
    NAME: CHRISTENBURY, LYNNE M.
    REGISTRATION NUMBER: 30,971
    REFERENCE/DOCKET NUMBER: CR-8926-C
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 302-992-5481
    TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1602 base pairs
    TYPE: Nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    ORGANISM: Glycine max
    STRAIN: Cultivar Wye
    DEVELOPMENTAL STAGE: Early seed fill
    HAPLOTYPE: Diploid
    TISSUE TYPE: Cotyledon
    ORGANELLE: Nucleus
  IMMEDIATE SOURCE:
    LIBRARY: cDNA to mRNA
    CLONE: 22B
  FEATURE:
    NAME/KEY: mat_peptide
    LOCATION: 271..1206
    IDENTIFICATION METHOD: Catalytically active
    IDENTIFICATION METHOD: when expressed in
    IDENTIFICATION METHOD: E. coli
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 106..1209
US-08-948-176-1

Query Match   0.6%; Score 21; DB 2; Length 1602;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3251 TTTTTATTTTTTAATAAAAA 3271
          |||||||||||||||||||
Db   87 TTTTTATTTTTTAATAAAAA 67

RESULT 12
PCT-US91-09160-1/c
; Sequence 1, Application PC/TUS9109160
; GENERAL INFORMATION:
  APPLICANT: Hitz, William D.
  APPLICANT: Yadav, Narendra S.
  TITLE OF INVENTION: Nucleotide Sequences of Soybean Acyl-ACP
  TITLE OF INVENTION: Thioesterase Genes
  NUMBER OF SEQUENCES: 22
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: E. I. du Pont de Nemours and Company
    STREET: 1007 Market Street
    CITY: Wilmington
    STATE: Delaware
    COUNTRY: U.S.A
    ZIP: 19898
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US91/09160
    FILING DATE: 19911216
    CLASSIFICATION:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/631,264
    FILING DATE: 20-DEC-1990
  ATTORNEY/AGENT INFORMATION:
    NAME: Morrissey, Bruce W.
    REGISTRATION NUMBER: 30,663
    REFERENCE/DOCKET NUMBER: CR-8926-A

```
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (302) 992-4927
;       TELEFAX:  (302) 892-7949
;       TELEX:  835420
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;          LENGTH:  1602 base pairs
;          TYPE:  NUCLEIC ACID
;          STRANDEDNESS:  single
;          TOPOLOGY:  linear
;       MOLECULE TYPE:  cDNA to mRNA
;       HYPOTHETICAL:  NO
;       ANTI-SENSE:  NO
;       ORIGINAL SOURCE:
;          ORGANISM:  Glycine max
;          STRAIN:  Cultivar Wye
;          DEVELOPMENTAL STAGE:  Early seed fill
;          HAPLOTYPE:  Diploid
;          TISSUE TYPE:  Cotyledon
;          ORGANELLE:  Nucleus
;       IMMEDIATE SOURCE:
;          LIBRARY:  cDNA to mRNA
;          CLONE:  22B
;       FEATURE:
;          NAME/KEY:  mat_peptide
;          LOCATION:  271..1206
;          IDENTIFICATION METHOD:  Catalytically active when
;          IDENTIFICATION METHOD:  expressed in E. coli
;       FEATURE:
;          NAME/KEY:  CDS
;          LOCATION:  106..1209
PCT-US91-09160-1

Query Match           0.6%;  Score 21;  DB 5;  Length 1602;
Best Local Similarity  100.0%;  Pred. No. 5.4;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3251 TTTTTATTTTTTAATAAAAA 3271
           |||||||||||||||||||
Db      87 TTTTTATTTTTTAATAAAAA 67

RESULT 13
US-10-204-708-80/c
;  Sequence 80, Application US/10204708
;  Patent No. 6677731
;  GENERAL INFORMATION:
;       APPLICANT:  OLEK, Alexander
;       APPLICANT:  PIEPENBROCK, Christian
;       APPLICANT:  BERLIN, Kurt
;       TITLE OF INVENTION:  Diagnosis of Diseases Associated with DNA Replication
;       TITLE OF INVENTION:  by Assessing DNA Methylation
;       FILE REFERENCE:  5013.1012
;       CURRENT APPLICATION NUMBER:  US/10/204,708
;       CURRENT FILING DATE:  2003-05-06
;       PRIOR APPLICATION NUMBER:  PCT/EP01/03971
;       PRIOR FILING DATE:  2001-04-06
;       PRIOR APPLICATION NUMBER:  DE 10019058.8
;       PRIOR FILING DATE:  2000-04-06
;       PRIOR APPLICATION NUMBER:  DE 10019173.8
;       PRIOR FILING DATE:  2000-04-07
;       PRIOR APPLICATION NUMBER:  DE 10032529.7
;       PRIOR FILING DATE:  2000-06-30
;       PRIOR APPLICATION NUMBER:  DE 10043826.1
;       PRIOR FILING DATE:  2000-09-01
;       NUMBER OF SEQ ID NOS:  98
;  SEQ ID NO 80
;       LENGTH:  8961
;       TYPE:  DNA
;       ORGANISM:  Artificial Sequence
;       FEATURE:
;       OTHER INFORMATION:  chemically treated genomic DNA (Homo sapiens)
;       FEATURE:
```

```
;       NAME/KEY:  unsure
;       LOCATION:  (5096)
;       OTHER INFORMATION:  n is a or g or c or t
US-10-204-708-80

Query Match           0.6%;  Score 21;  DB 4;  Length 8961;
Best Local Similarity  100.0%;  Pred. No. 5.5;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3251 TTTTTATTTTTTAATAAAAA 3271
           |||||||||||||||||||
Db    1629 TTTTTATTTTTTAATAAAAA 1609

RESULT 14
US-09-134-000C-2872
;  Sequence 2872, Application US/09134000C
;  Patent No. 6617156
;  GENERAL INFORMATION:
;       APPLICANT:  Lynn Doucette-Stamm et al
;       TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;       TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;       FILE REFERENCE:  032796-032
;       CURRENT APPLICATION NUMBER:  US/09/134,000C
;       CURRENT FILING DATE:  1998-08-13
;       PRIOR APPLICATION NUMBER:  US 60/055,778
;       PRIOR FILING DATE:  1997-08-15
;       NUMBER OF SEQ ID NOS:  6812
;       SOFTWARE:  PatentIn version 3.1
;  SEQ ID NO 2872
;       LENGTH:  591
;       TYPE:  DNA
;       ORGANISM:  Enterococcus faecalis
US-09-134-000C-2872

Query Match           0.6%;  Score 20;  DB 4;  Length 591;
Best Local Similarity  100.0%;  Pred. No. 16;
Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2385 ATCAATTATTATTTTTTTTT 2404
           ||||||||||||||||||||
Db     410 ATCAATTATTATTTTTTTTT 429

RESULT 15
US-08-585-593A-27/c
;  Sequence 27, Application US/08585593A
;  Patent No. 6503706
;  GENERAL INFORMATION:
;       APPLICANT:  ABKEN, Hinrich J
;       APPLICANT:  ALBERT, Winfried
;       APPLICANT:  JUNGFER, Herbert
;       TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
;       TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
;       TITLE OF INVENTION:  FORMATION
;       NUMBER OF SEQUENCES:  66
;       CORRESPONDENCE ADDRESS:
;          ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;          STREET:  655 Fifteenth Street N.W. Suite 330
;          CITY:  Washington
;          STATE:  D.C.
;          COUNTRY:  U.S.A.
;          ZIP:  20005-5701
;       COMPUTER READABLE FORM:
;          MEDIUM TYPE:  Floppy disk
;          COMPUTER:  IBM PC compatible
;          OPERATING SYSTEM:  PC-DOS/MS-DOS
;          SOFTWARE:  PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;          APPLICATION NUMBER:  US/08/585,593A
;          FILING DATE:  16-JAN-1996
;          CLASSIFICATION:  435
;       PRIOR APPLICATION DATA:
```

```
    APPLICATION NUMBER:  PCT/EP94/02307
;    FILING DATE: 13-JUL-1994
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  DE P 43 23 727.4
;    FILING DATE: 15-JUL-1993
;  ATTORNEY/AGENT INFORMATION:
;    NAME: Murray, Robert B.
;    REGISTRATION NUMBER: 22,980
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:   (202)638-5000
;    TELEFAX:   (202)638-4810
;  INFORMATION FOR SEQ ID NO:  27:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 592 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS:  single
;    TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-585-593A-27

Query Match             0.6%;  Score 20;  DB 4;  Length 592;
Best Local Similarity  100.0%;  Pred. No. 16;
Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    3253 TTTATTTTTTTAATAAAAAT 3272
           |||||||||||||||||||
Db     109 TTTATTTTTTTAATAAAAAT 90


Search completed: July 21, 2004, 19:27:57
Job time : 252 secs
```

OM nucleic - nucleic search, using sw model

```
        GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:        July 21, 2004, 17:06:46 ; Search time 1433 Seconds
                                          (without alignments)
                            11616.099 Million cell updates/sec

Title:          US-09-270-437D-6
Perfect score:  3412
Sequence:       1 ggcagcggaggaggcgagga.........aacctgaaatgttatt 3412

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Word size :     0

Searched:       3191023 seqs, 2433312756 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Total number of hits satisfying chosen parameters:    6382046

Post-processing: Listing first 45 summaries

Database :      Published_Applications_NA.*
```

```
 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Str	Score	% Query Match	Length	DB	ID	Description
1		3412	100.0	3412	9	US-09-899-651-6	Sequence 6, Appli
2		2953	86.5	3694	9	US-09-764-864-329	Sequence 329, App
3		2143	62.8	3283	9	US-09-899-651-8	Sequence 8, Appli
4		1861	54.5	3905	10	US-09-814-353-20241	Sequence 20241, A
5		1810	53.0	3667	15	US-10-097-340-146	Sequence 146, App
6		1810	53.0	3667	17	US-10-648-593-48	Sequence 48, Appl
7		1496	43.8	2290	9	US-09-764-864-332	Sequence 332, App
8		401	11.8	1186	9	US-09-764-864-749	Sequence 749, App
9		397	11.6	582	10	US-09-814-353-14725	Sequence 14725, A
10		389	11.4	822	9	US-09-764-864-330	Sequence 330, App
11		347	10.2	579	10	US-09-814-353-1995	Sequence 1995, Ap
12		347	10.2	579	10	US-09-814-353-8341	Sequence 8341, Ap
13		333	9.8	472	15	US-10-066-543-2949	Sequence 2949, Ap
14		264	7.7	264	9	US-09-777-564-1124	Sequence 1124, Ap
15		264	7.7	264	15	US-10-015-219-1124	Sequence 1124, Ap
16	c	263	7.7	263	9	US-09-777-564-1307	Sequence 1307, Ap
17		263	7.7	263	9	US-09-777-564-1537	Sequence 1537, Ap
18	c	263	7.7	263	15	US-10-015-219-1307	Sequence 1307, Ap
19	c	263	7.7	263	15	US-10-015-219-1537	Sequence 1537, Ap
20		222	6.5	222	9	US-09-815-34-676	Sequence 676, App
21		222	6.5	222	9	US-09-815-343-1285	Sequence 1285, Ap
22		222	6.5	222	13	US-10-097-105-676	Sequence 676, App
23		222	6.5	222	13	US-10-097-105-1285	Sequence 1285, Ap
24		184	5.4	531	15	US-10-066-543-1888	Sequence 1888, Ap
25		167	4.9	218	9	US-09-815-343-1415	Sequence 1415, Ap
26		167	4.9	218	13	US-10-097-105-1415	Sequence 1415, Ap
27		165	4.8	500	9	US-09-783-590-1246	Sequence 1246, Ap
28		143	4.2	402	10	US-09-918-995-5484	Sequence 5484, Ap
29		140	4.1	243	13	US-09-815-343-614	Sequence 614, App
30		140	4.1	243	16	US-10-097-105-614	Sequence 614, App
31		135	4.0	1707	9	US-10-262-445-39	Sequence 39, Appl
32		93	2.7	253	13	US-09-815-343-729	Sequence 729, App
33		93	2.7	253	9	US-10-097-105-729	Sequence 729, App
34		91	2.7	402	9	US-09-783-590-5490	Sequence 5490, Ap
35	c	83	2.4	380	9	US-09-964-824A-349	Sequence 349, App
36	c	83	2.4	380	9	US-09-880-107-1135	Sequence 1135, Ap
37	c	81	2.4	590	13	US-10-240-425-760	Sequence 760, App
38		81	2.4	161484	17	US-10-317-401-4	Sequence 4, Appli
39	c	68	2.0	412	9	US-09-962-436-95	Sequence 95, Appl
40		62	1.8	145	9	US-09-764-866-747	Sequence 747, App
41		55	1.6	399	9	US-10-029-386-14468	Sequence 14468, A
42	c	55	1.6	567	15	US-10-029-386-765	Sequence 765, App
43	c	50	1.5	154	15	US-09-783-590-5140	Sequence 5140, Ap
44		37	1.1	495	9	US-09-918-995-26330	Sequence 26330, A
45	c	26	0.8	26	10	US-10-262-445-95	Sequence 95, Appl

ALIGNMENTS

```
RESULT 1
US-09-899-651-6
; Sequence 6, Application US/0989651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-6

  Query Match             100.0%;  Score 3412;  DB 9;  Length 3412;
  Best Local Similarity   100.0%;  Pred. No.0;
  Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 GGCAGCGGAGGAGGCGAGGAGCGGCGGAGGAGCCCGGGTACCCGGCGGGGAGCCGGGTCTCGGGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCGGAGGAGGCGAGGAGCGGCGGAGGAGCCCGGGTACCCGGCGGGGAGCCGGGTCTCGGGG 60

Qy   61 AAGAGACCGGATGATGATGAACAAGCTTTACATCGGGAACGTTTACATCGGGAACCTGAGCCCGCGCCGTCACCGCCGAC 120
```

```
Db        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   61     AAGAGACGGATGATGAACAAGCTTTACATCGGGAAGAACCTGAGCCCGCCCGCCCACCGAC   120

Qy  121   GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCTGGGACAGTCCTGCTGAAG   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121   GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCTGGGACAGTCCTGCTGAAG   180

Qy  181   TCCGGCTACGCCTTCGTGTGACTACCCGACCAGAACTGGGCCATCCGGCCATCGAGACC   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181   TCCGGCTACGCCTTCGTGTGACTACCCGACCAGAACTGGGCCATCCGGCCATCGAGACC   240

Qy  241   CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241   CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA   300

Qy  301   AAGTCAAGGAGCAGGAAAAATTCAGATTCGAAACATCCCTCCCAGCTGGGACAGTGGGAGTG   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301   AAGTCAAGGAGCAGGAAAAATTCAGATTCGAAACATCCCTCCCAGCTGGGACAGTGGGAGTG   360

Qy  361   TTGGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGTGGAACAAGTCAACACAGAC   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361   TTGGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGTGGAACAAGTCAACACAGAC   420

Qy  421   ACAGAAACCGCCGTGTCAACGTCACATATGCAACAAGGAAGCAAAAATAGCCATG   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421   ACAGAAACCGCCGTGTCAACGTCACATATGCAACAAGGAAGCAAAAATAGCCATG   480

Qy  481   GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCCTTCAAGATTCCTACATCCCGGAT   540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481   GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCCTTCAAGATTCCTACATCCCGGAT   540

Qy  541   GAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGTGGGGACCACTCTTCCCGG   600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541   GAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGTGGGGACCACTCTTCCCGG   600

Qy  601   GAGCAAGGCCACGCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCGG   660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601   GAGCAAGGCCACGCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCGG   660

Qy  661   ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG   720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661   ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG   720

Qy  721   AACATCACTAAGGACAGAGACCCAGTCCCGGTAGTATCGGGTAGATATCCATAGAAAGAAGAACTCTGGAGCT   780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721   AACATCACTAAGGACAGAGACCCAGTCCCGGTAGTATCGGGTAGATATCCATAGAAAGAAGAACTCTGGAGCT   780

Qy  781   GCAGAGAAGCCTGTCACCATCATGCCATGGACTTCTGAAGCATGCCGCATG   840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781   GCAGAGAAGCCTGTCACCATCATGCCATGGACTTCTGAAGCATGCCGCATG   840

Qy  841   ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCTG   900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841   ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCTG   900

Qy  901   AAAATCTTGCACACAACAATGGCTTGGTTGAAGACTGATTGGAAAGAAGGCAGAAATTTG   960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901   AAAATCTTGCACACAACAATGGCTTGGTTGAAGACTGATTGGAAAGAAGGCAGAAATTTG   960

Qy  961   AAGAAAATGAAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC   1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961   AAGAAAATGAAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC   1020

Qy  1021  ATATACACCCGGAAGAAGACCATCACTGTGAAGGGCACAGTTGAGGCCTTGAGCGCTGCT   1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  ATATACACCCGGAAGAAGACCATCACTGTGAAGGGCACAGTTGAGGCCTTGAGCGCTGCT   1080

Qy  1081  GAGATAGAGATTATGAAGAAGTCGCTGCGTAAATGATATGCTGGCTGTTAAC   1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GAGATAGAGATTATGAAGAAGTCGCTGCGTAAATGATATGCTGGCTGTTAAC   1140

Qy  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCCGACTTGCATCTTTTCAACAAGGA   1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCCGACTTGCATCTTTTCAACAAGGA   1200

Qy  1201  CTGTCCGTGCTATCTCCACCACCAGCAGGGCCCCCGCGGAGCTCCCCGTGCCCCGTCACCAC   1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CTGTCCGTGCTATCTCCACCACCAGCAGGGCCCCCGCGGAGCTCCCCGTGCCCCGTCACCAC   1260

Qy  1261  CCCTTCACTCACCACTCGGATACTTCTCCAGCAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCACAG   1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  CCCTTCACTCACCACTCGGATACTTCTCCAGCAGCAGGAGAGATTGTGAATCTCTTCATCCCAACCACAG   1320

Qy  1321  TTCCCGATCATCACTCTTATCCAGACCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAG   1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  TTCCCGATCATCACTCTTATCCAGACCAGGAGAGATTGTGAATCTCTTCATCCCAACCCAG   1380

Qy  1381  GCTGTGGGCGCCATCATCGGGAAGAAGGGGGCACACATCAAAACAGCTGGGCGAGATTCGCC   1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  GCTGTGGGCGCCATCATCGGGAAGAAGGGGGCACACATCAAAACAGCTGGGCGAGATTCGCC   1440

Qy  1441  GGAGCCTCTATCAAGATTGCCCCTGCCCTGGGAAGCGTCAGCGAAAGGATGGTCATC   1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  GGAGCCTCTATCAAGATTGCCCCTGCCCTGGGAAGCGTCAGCGAAAGGATGGTCATC   1500

Qy  1501  ATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA   1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  ATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA   1560

Qy  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAGCTGAAGCGCATATCAGAGTGCCC   1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCGCATATCAGAGTGCCC   1620

Qy  1621  TCTTCCACAGCTGCCGGGTGATTGGCAAAGTGGCAAGGACCGTGAACGAACTGCAGAAC   1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  TCTTCCACAGCTGCCGGGTGATTGGCAAAGTGGCAAGGACCGTGAACGAACTGCAGAAC   1680

Qy  1681  TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAGCCCAGATGAAAATGAGGAAGTG   1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAGCCCAGATGAAAATGAGGAAGTG   1740

Qy  1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGAGCCAGAAATACCCTCAGGGAA   1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGAGCCAGAAATACCCTCAGGGAA   1800

Qy  1801  ATTGTACAACAGGTGAAGCAGCAGCAGGCACCAGCAGAAATACCCTCACAGCGC   1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  ATTGTACAACAGGTGAAGCAGCAGCAGGCACCAGCAGAAATACCCTCACAGCGC   1860

Qy  1861  AGCAAGTGAGGCTCCCACAGGCACCGATGAATGTAGCCCTTCCAACAC   1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  AGCAAGTGAGGCTCCCACAGGCACCGATGAATGTAGCCCTTCCAACAC   1920

Qy  1921  CTGACAGAATGAGACCAAACGCCAGCCAGGGACCTCGCCAGGAGACCATCTGAG   1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  CTGACAGAATGAGACCAAACGCCAGCCAGGGACCTCGCCAGGAGACCATCTGAG   1980

Qy  1981  GAATGAGAGTCTGCGGAGAAGGTCAGCCAGGTTTGCCAGAACCCCAGGGGC   2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  GAATGAGAGTCTGCGGAGAAGGTCAGCCAGGTTTGCCAGAACCCCAGGGGC   2040

Qy  2041  CGGAGGGGGCGGGAAGGTCAGCGCCATCCAGCGGATCTCCACTCACTCGGATCTCCTGAA   2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGGAGGGGGCGGGAAGGTCAGCGCCATCCAGCGGATCTCCACTCACTCGGATCTCCTGAA   2100

Qy  2101  CCCCAGGGCTTCTGCAGGCTTCAGCGCTATCCTTTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGC   2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CCCCAGGGCTTCTGCAGGCTTCAGCGCTATCCTTTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGC   2160

Qy  2161  CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCAAGC   2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCAAGC   2220

Qy  2221  AAAATGCACACCCCTTTTCTGTGGCAAATGCTGTCTCTGTACATGTGTGTACATATTAGAAA   2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AAAATGCACACCCCTTTTCTGTGGCAAATGCTGTCTCTGTACATGTGTGTACATATTAGAAA   2280
```

```
Qy   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT  3412
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT  3412
```

RESULT 2
US-09-764-864-329
; Sequence 329, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 329
;   LENGTH: 3694
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-329

Query Match   86.5%;   Score 2953;   DB 9;   Length 3694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 3303;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   2;

```
Qy    68  GGATGATGAACAAGCTTACATCGGACCTTACATCGGACCTTCACCCGCGTCACCCGCGACCTCC  127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91  GGATGATGAACAAGCTTACATCGGACCTTACATCGGACCTTCACCCGCGTCACCCGCGACCTCC  150

Qy   128  GGCAGCTCTTTGGGACAGGAGCTGCCCCTGCGGGACAGTCCTGCTGAAGTCCGGCT  187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151  GGCAGCTCTTTGGGACAGGAGCTGCCCCTGCGGGACAGTCCTGCTGAAGTCCGGCT  210

Qy   188  ACGCCTTCGTGGACTACCCCGACCAGACTGGGCCATCCGCGGCCATCGAGACCCTCTGG  247
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211  ACGCCTTCGTGGACTACCCCGACCAGACTGGGCCATCCGCGGCCATCGAGACCCTCTGG  270

Qy   248  GTAAAGTGGAATTGCATGGGAAAATCATGGAAAATCATGGAGTTGATTACTCAGTCTCTAAAAAGCTAA  307
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271  GTAAAGTGGAATTGCATGGGAAAATCATGGAAAATCATGGAGTTGATTACTCAGTCTCTAAAAAGCTAA  330

Qy   308  GGAGCAGCAGAAATTCAGATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGATG  367
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331  GGAGCAGCAGAAATTCAGATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGATG  390

Qy   368  GACTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAA  427
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391  GACTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAA  450

Qy   428  CCGCCGTTGTCAACGTCACATATGCAACAAGACAGAAGCAAAAATAGCCATGGAGAAGC  487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451  CCGCCGTTGTCAACGTCACATATGCAACAAGACAGAAGCAAAAATAGCCATGGAGAAGC  510

Qy   488  TAAGCGGGGCATCAGTTGGAACTACTCCTTGAGAACTCCTTCCTACATCCCGGATGAAGAGG  547
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511  TAAGCGGGGCATCAGTTGGAACTACTCCTTGAGAACTCCTTCCTACATCCCGGATGAAGAGG  570

Qy   548  TGAGCTCCCCTTCGCCCCCCAGCGTGGGACCCACTCTTCCCGGGACCAAG  607
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   571  TGAGCTCCCCTTCGCCCCCCAGCGTGGGACCCACTCTTCCCGGGACCAAG  630

Qy   608  GCCACGCCCCTGGGGGCACTTCTCAGGGCCGCAGAGATTGAATTCCCGTCCGATCCTCTG  667
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631  GCCACGCCCCTGGGGGCACTTCTCAGGGCCGCAGAGATTGAATTCCCGTCCGATCCTCTG  690

Qy   668  TCCCACCCAGTTGTTGGTGCCATCATGGAAAGGAGGGCTTGACCATAAGAACATCA  727
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   691  TCCCACCCAGTTGTTGGTGCCATCATGGAAAGGAGGGCTTGACCATAAGAACATCA  750

Qy   728  CTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCGAGA  787
```

```
Qy   2281  GGGAAGATGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT  2340
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  GGGAAGATGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGCGGTAATATAT  2340

Qy   2341  TTTAGAAATAATATATCAAATAACTCCAATTTTAATCAATTATTAATTT  2400
               |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTT  2400

Qy   2401  TTTTTCTTTTAAAGAGAAGAACAGCCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  TTTTTCTTTTAAAGAGAAGAACAGCCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460

Qy   2461  GTCTCACGGGTGTAGAGAGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT  2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461  GTCTCACGGGTGTAGAGAGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGAAAT  2520

Qy   2521  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA  2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA  2580

Qy   2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCCATGCTCGTGTCTAGCTCATCATACACCCATT  2640
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCCATGCTCGTGTCTAGCTCATCATACACCCATT  2640

Qy   2641  TCTCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT  2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641  TCTCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCT  2700

Qy   2701  CTCTCTGTTTATCTCTCCCCTTCTTCTCCATCTCCATTCTTTTGAA  2760
               ||||||||||||||||||||||||||||||||||||||||||||||
Db   2701  CTCTCTGTTTATCTCTCCCCTTCTTCTCCATCTCCATTCTTTTGAA  2760

Qy   2761  TTTTCTCATCCCTCCATCTCAATCCGTATCTACGGCACCACCCCCCCCCCAGCAAAGCA  2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761  TTTTCTCATCCCTCCATCTCAATCCGTATCTACGGCACCACCCCCCCCCCAGCAAAGCA  2820

Qy   2821  GTGCTCTGAGTATCACATCACACAGAACAAAAGCGAAACCACAACACCAGCCTCAA  2880
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  GTGCTCTGAGTATCACATCACACAGAACAAAAGCGAAACCACAACACCAGCCTCAA  2880

Qy   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGGAAG  2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGGAAG  2940

Qy   2941  AGGAAAACAGGAACCCACCAAACCACCAATCAACCAATCAACCAAGAAAAATTCCACAATG  3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  AGGAAAACAGGAACCCACCAAACCACCAATCAACCAATCAACCAAGAAAAATTCCACAATG  3000

Qy   3001  AAAGAATGTATTTTGTCTTTTGTGTATAAGCCATCAATATTCAGCAAAATG  3060
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  AAAGAATGTATTTTGTCTTTTGTGTATAAGCCATCAATATTCAGCAAAATG  3060

Qy   3061  ATTTCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGTTGGCC  3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  ATTTCTTTCTTTAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGTTGGCC  3120

Qy   3121  CAGGGGGTTAAATTCACAGATTTTTTCCCTTGTGTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT  3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  CAGGGGGTTAAATTCACAGATTTTTTCCCTTGTGTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT  3180

Qy   3181  GTGTTTTTACCTCAGCACCTTGCTCTTGTCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTCGATAG  3240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  GTGTTTTTACCTCAGCACCTTGCTCTTGTCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTCGATAG  3240

Qy   3241  TTGGAGCATTTTTTATTTTTATTTTTAATAAAATGAGTTGGAAAATGAGTTGGAAAAATAAGATATCAACT  3300
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241  TTTGGAGCATTTTTTATTTTTATTTTTAATAAAATGAGTTGGAAAATGAGTTGGAAAAATAAGATATCAACT  3300

Qy   3301  GCCAGCCTGGAGAAGGTGACAGTGCAACAGCTGTTCGCAACAGTGTCTTCCCGCT  3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  GCCAGCCCAGTTGTTGGTGCAACAGTGTCGCAACAGTGTCTTCTGAAATTGTCTTCTGAAATTGTCTTCCGCT  3360
```

```
Db  1831 AACAGGTGAAGCAGCAGGAGCAGCAGGAGCAGCAGAAATACCCTCAGGAGTCGCCTCACAGCGGCAGCAAGT 1890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1868 GAGGCTCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAG 1927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1891 GAGGCTCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAG 1950

Qy  1928 AATGAGACCAAACGCAGCCAGCCAGATCGGAGGCAAACCAATCCAGGAGGATGAG 1987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1951 AATGAGACCAAACGCAGCCAGCCAGATCGGAGGCAAACCAATCCAGGAGGATGAG 2010

Qy  1988 AAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCTGAGAACCCGAGGGCCGAGGAG 2047
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2011 AAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCTGAGAACCCGAGGGCCGAGGAG 2070

Qy  2048 GGGCGGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCGCCTCCCCGCCCCCAGG 2107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2071 GGGCGGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCGCCTCCCCGCCCCCAGG 2130

Qy  2108 GCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCACTCCTGGATCTCTCCTGAACTCCCAC 2167
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2131 GCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCACTCCTGGATCTCTCCTGAACTCCCAC 2190

Qy  2168 GACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2191 GACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2250

Qy  2228 ACACCCTTTTTCTGTCGTAAATCGTCTCTGTACATCGTCTCTGTACATATTAGAAAGGGAAGA 2287
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2251 ACACCCTTTTTCTGTCGTAAATCGTCTCTGTACATCGTCTCTGTACATATTAGAAAGGGAAGA 2310

Qy  2288 TGTTAAGATATGTGGCGTGTGGTTACACAGGGTGCCTGCGAGCGGTAATATTTTAGAA 2347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2311 TGTTAAGATATGTGGCGTGTGGTTACACAGGGTGCCTGCGAGCGGTAATATATTTTAGAA 2370

Qy  2348 ATAATATATCAAATACTCAACTAACTCCAATTTTAATCAATTTTAATCAATTATTAATTTTTTTCT 2407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2371 ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTATTTTTTTCT 2430

Qy  2408 TTTTAAGAGAAAGCAGGCTTTGAGGCCACCCGACAAAATTCACCCGAGAGGAAATCTCGTCG 2467
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2431 TTTTAAGAGAAAGCAGGCTTTGAGGCCACCCGACAAAATTCACCCGAGAGGAAATCTCGTCG 2490

Qy  2468 GGTGTAGAGAGGAGCTTTGAGGCTTGAGGCCACCCGATCAACAGGATCATCGTCGTCG 2527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2491 GGTGTAGAGAGGAGCTTTGAGGCTTGAGGCCACCCGATCAACAGGATCATCGTCGTCG 2550

Qy  2528 GAAGGACACTCACGGCAGTTCTGGATCAACAGCGTTGTATGTCAACAGAAGGGATACCGTCTC 2587
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2551 GAAGGACACTCACGGCAGTTCTGGATCAACAGCGTTGTATGTCAACAGAAGGGATACCGTCTC 2610

Qy  2588 CTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATCACCCATTTCTCTTT 2647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2611 CTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATCACCCATTTCTCTTT 2670

Qy  2648 GCTTCACAGGTTTTAAACTCGTTTTTGCATACTGCTATATAATTCTCTGTCTATAATTCTCTGTCTCTCTG 2707
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2671 GCTTCACAGGTTTTAAACTCGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTG 2730

Qy  2708 TTTATCTCTCCCCTCCCCTCCCCTCCTTCTTCTCCATTCTTTGAATTCCTC 2767
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2731 TTTATCTCTCCCCTCCCCTCCCCTCCTTCTTCTCCATTCTTTGAATTCCTC 2790

Qy  2768 ATCCCTCCATCTCAATCCGTATCTACGA--CCCCCCCCCCCAGGCAAAGCAGTGCT 2825
              ||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db  2791 ATCCCTCCATCTCAATCCGTATCTCACGCACCCCCCCCCCCCAGGCAAAGCAGTGCT 2850

Qy  2826 CTGAGTATCACATCACAAAAGAACAAAAGCGAAACAACAAACCAGCCTCAACTTAC 2885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2851 CTGAGTATCACATCACACAAAAGAACAAAAGCGAAACAACAAACCAGCCTCAACTTAC 2910

Qy  2886 ACTTGGTTACTCAAAAGAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2911 ACTTGGTTACTCAAAAGAACAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAGAGGAA 2970
```

```
Db   751 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         CTAAGCAGACCCAGTCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTCGAGACGA 810

Qy   788 AGCCTTGTCACCATCCATGCCAGAGGGACTTCTGAAGCATGCGCCATGATTCTTG 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   811 AGCCTTGTCACCATCCATGCCAGAGGGACTTCTGAAGCATGCGCCATGATTCTTG 870

Qy   848 AAATCATGCAGAAAGAGACGATGAGAACTAGCCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   871 AAATCATGCAGAAAGAGACGATGAGAACTAGCCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 930

Qy   908 TGGCACACAATGGCTTGGTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATTTGAAGAAAA 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   931 TGGCACACAATGGCTTGGTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATTTGAAGAAAA 990

Qy   968 TTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCATATACA 1027
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   991 TTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCATATACA 1050

Qy  1028 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAG 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1051 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAG 1110

Qy  1088 AGATTATGAAGAAGTGCGTGAGGGCCTTTGAAAAATGATATGCTGGCTGTTAACCAACAG 1147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1111 AGATTATGAAGAAGTGCGTGAGGGCCTTTGAAAAATGATATGCTGGCTGTTAACCAACAAG 1170

Qy  1148 CCAATCTGATCCCAGGGTTGAACCTCAGCGCATCTTTTCAACAGGACTGTTCCG 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1171 CCAATCTGATCCCAGGGTTGAACCTCAGCGCATCTTTTCAACAGGACTGTTCCG 1230

Qy  1208 TGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCGTGCCCCCTACCACCCCTTCA 1267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1231 TGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCGTGCCCCCTACCACCCCCTTCA 1290

Qy  1268 CTACCCACTCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGCCCGTTCGCCGC 1327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1291 CTACCCACTCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGCCCGTTCGCCGC 1350

Qy  1328 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAAACAGCTGGTGTGG 1387
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1351 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAAACAGCTGGTGTGG 1410

Qy  1388 GCGCCATCATCGGAGAAGAAGGGGCACACATCAAACAGCTGGCGAAGGCCT 1447
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1411 GCGCCATCATCGGAGAAGAAGGGGGCACACATCAAACAGCTGGCGAAGGCCT 1470

Qy  1448 CTATCAAGATTGCCCCTCGCGGAAGGCCCAGACGTTGGCAAGGATGGTCATCATCACCG 1507
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1471 CTATCAAGATTGCCCCTCGCGGAAGGCCCAGACGTGCAAGGATGGTCATCATCACCG 1530

Qy  1508 GGCCACCCGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAAGAGAGAAA 1567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1531 GGCCACCCGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAAGAGAGAAA 1590

Qy  1568 ACTTCTTTAACCCAAAGAAGAGAGTGAAGCTCATATCAGATGCCCTCTTCCA 1627
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1591 ACTTCTTTAACCCAAAGAAGAGAGTGAAGCTCGAATCAGATGCCCTCTTCCA 1650

Qy  1628 CAGCTGGCGGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCA 1687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1651 CAGCTGGCGGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCA 1710

Qy  1688 GTGCCAGAAGTCATCGTGCCGTGACCAAACCGCAGATGAAAATGAGGAAGTGATCGTCA 1747
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1711 GTGCAGAAGTCATCGTGCCTCGTGACCAAACCGCAGATGAAAATGAGGAAGTGATCGTCA 1770

Qy  1748 GAATTATCGGGCACTTCTTTGCTTAGCCAGACTGCAAGATCAGAACTTGTAC 1807
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1771 GAATTATCGGGCACTTCTTTGCTTAGCCAGACTGCAAGATCAGAACTTGTAC 1830

Qy  1808 AACAGGTGAAGCAGCAGGAGCAGCAGGAGCAGCAGAAATACCCTCAGGAGTCGCCTCACAGCGGCAGCAAGT 1867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  1390  GCCATCATCGGGAAGAGGAGGGGGCACACATCAAACATCAAACAGCTGGCGGAGAATTCGCCGGAGCTCT  1449
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  GCCATCATCGGGAAGAGGAGGGGGCACACATCAAACATCAAACAGCTGGCGGAGAATTCGCCGGAGCTCT  1320

Qy  1450  ATCAAGATGCCCCTGGGAAGCCCAGACGTTCAGCGAAAGATGGTCATCATCACCGGG  1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  ATCAAGATGCCCCTGGGAAGCCCAGACGTTCAGCGAAAGATGGTCATCATCACCGGG  1380

Qy  1510  CCACCGGAAGCCCAGTTCAAGGCCCAGGTTCAAGGCCGATCTTTGGGAAACTGAAAGAGGAAAAC  1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  CCACCGGAAGCCCAGTTCAAGGCCCAGGTTCAAGGCCGATCTTTGGGAAACTGAAAGAGGAAAAC  1440

Qy  1570  TTCTTTAACCCAAAGAAGAAGTGAAGCGCATATCAGAGTGCCCTTCCACA  1629
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  TTCTTTAACCCAAAGAAGAAGTGAAGCGCATATCAGAGTGCCCTTCCACA  1500

Qy  1630  GCTGGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGAACTTAACCAGT  1689
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  GCTGGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGAACTTAACCAGT  1560

Qy  1690  AGCAAGTCATCGTGCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA  1749
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  AGCAAGTCATCGTGCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTGATCGTCAGA  1620

Qy  1750  ATTATCGGGCACTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA  1809
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  ATTATCGGGCACTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA  1680

Qy  1810  CAGGTGAAGCAGCAGGAGCACCAGCAGAAATACCCTCAGGGAGTCGCTCACAGCGCAAGTGA  1869
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  CAGGTGAAGCAGCAGGAGCACCAGCAGAAATACCCTCAGGGAGTCGCTCACAGCGCAAGTGA  1740

Qy  1870  GGCTCCCACAGGGCACCAGCAGCAAAACAACGGATGAATGAATGAGCCCTTCCAACACTGAGGAAGAA  1929
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  GGCTCCCACAGGGCACCAGCAGCAAAACAACGGATGAATGAATGAGCCCTTCCAACACTGAGGAAGAA  1800

Qy  1930  TGAGACCAAAGCAGCGAGCCAGCCAGCGGAGCAAAACCAAAAGACCATCTGAGGAATGAGAA  1989
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  TGAGACCAAAGCAGCGAGCCAGCCAGCGGAGCAAAACCAAAAGACCATCTGAGGAATGAGAA  1860

Qy  1990  GTCTGCGGAGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCCAGGAGAGGG  2049
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  GTCTGCGGAGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCCAGGAGAGGG  1920

Qy  2050  GCGGGGAAGGTCAGCCAGGTTGCCAGGAACCACCAGGACGCCCGCCTCCCGCCCCCCAGGGC  2109
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  GCGGGGAAGGTCAGCCAGGTTGCCAGGAACCACCAGGACGCCCGCCTCCCGCCCCCCAGGGC  1980

Qy  2110  TTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCCACGA  2169
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  TTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTCTCCTGAACTCCCACGA  2040

Qy  2170  CGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2100

Qy  2230  ACCCTTTTCTGTGTGGCAAATCGTCTCTGTACATTAGAAAGGGAAGATG  2289
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  ACCCTTTTCTGTGTGGCAAATCGTCTCTGTACATTAGAAAGGGAAGATG  2160

Qy  2290  TTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCGGGTAATATATTTTAGAAAT  2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  TTAAGATATGTGGCCTGTGTGGGTTACACAGGGTGCCTGCAGCGGGTAATATATTTTAGAAAT  2220

Qy  2350  AATATATCAAATAACTAACTAACTAACTCCAATTTTAATCAATTTTAATCAATTATTAATTTTTTTCTTT  2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AATATATCAAATAACTAACTAACTAACTCCAATTTTAATCAATTTTAATCAATTATTAATTTTTTTCTTT  2280

Qy  2410  TTAAAGAGAAAGCAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGGGC  2469
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  TTAAAGAGAAAGCAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGGGC  2340
```

```
Qy  2946  AACAGGAACCCACCAAAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGA  3005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2971  AACAGGAACCCACCAAAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGA  3030

Qy  3006  ATGTATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC  3065
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3031  ATGTATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC  3090

Qy  3066  TTTCTTT-AAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG  3124
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3091  TTTCTTTAAAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG  3150

Qy  3125  GCGTTAAATTCACAGATTTTTTAACGAGAAAACACAGAGAGAAGCTACCTCAGTGT  3184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3151  GCGTTAAATTCACAGATTTTTTAACGAGAAAACACAGAGAGAAGCTACCTCAGTGT  3210

Qy  3185  TTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3211  TTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3270

Qy  3245  AGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA  3304
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3271  AGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA  3330

Qy  3305  GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC  3364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3331  GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC  3390

Qy  3365  AAGAACC  3371
          |||||||
Db  3391  AAGAACC  3397
```

RESULT 3
US-09-899-651-8

; Sequence 8, Application US/09899651
; Patent No. US2002011470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-8

Query Match   62.8%;   Score 2143; DB 9;   Length 3283;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2143; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy  1270  ACCCACTCCGGATACTTCTCCGAGCAGGAGATTGTCAACTCCCATCACCAGTTGGCCCGTTCCCCGCAT  1329
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  ACCCACTCCGGATACTTCTCCGAGCAGGAGATTGTCAACTCCCATCACCAGTTGGCCCGTTCCCCGCAT  1200

Qy  1330  CATCACTCTTATCCAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGC  1389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CATCACTCTTATCCAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGC  1260
```

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 20241
LENGTH: 3905
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20241

Query Match   54.5%;   Score 1861;   DB 10;   Length 3905;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 2101;   Conservative   0;   Mismatches   0;   Indels   2;   Gaps   2;

```
Qy  1270  ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTCCCGCAT  1329
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1507  ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTCCCGCAT  1566

Qy  1330  CATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGC  1389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1567  CATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGC  1626

Qy  1390  GCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGGCGGAGCCTCT  1449
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1627  GCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGGCGGAGCCTCT  1686

Qy  1450  ATCAAGATTGCCCCTCGGGAAGGCCCAGAGCGTCAGGGAAAGGATGTCATCATCACCGGG  1509
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1687  ATCAAGATTGCCCCTCGGGAAGGCCCAGAGCGTCAGGGAAAGGATGTCATCATCACCGGG  1746

Qy  1510  CCACCGGAAGCCAGTTCAAGGCAGGGACGGATCTTTGGGAAACTGAAGAGAGGAAAAC  1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1747  CCACCGGAAGCCAGTTCAAGGCAGGGACGGATCTTTGGGAAACTGAAGAGAGGAAAAC  1806

Qy  1570  TTCTTTAACCCAAAGAAGAGAGTGAAGCTGAAGGCGCATATCAGAGTGCCCTTCCACA  1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1807  TTCTTTAACCCAAAGAAGAGAGTGAAGCTGAAGGCGCATATCAGAGTGCCCTTCCACA  1866

Qy  1630  GCTGGCCGGGTGATTGGCAAGACCGTCAACGAACCGTGAACAACTGGCAAGAACCTTAACCAGT  1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1867  GCTGGCCGGGTGATTGGCAAGACCGTCAACGAACCGTGAACAACTGGCAAGAACCTTAACCAGT  1926

Qy  1690  GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGA  1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1927  GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGA  1986

Qy  1750  ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAAATTGTACAA  1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1987  ATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGAAATTGTACAA  2046

Qy  1810  CAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAGTGA  1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2047  CAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCAGCAAGTGA  2106
```

```
Qy  2470  TGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2529
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2400

Qy  2530  AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACCGTCTCCT  2589
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACCGTCTCCT  2460

Qy  2590  TGAAGAGGAAACTCTGTCACTCCTAGCTCATACACCCATTCTCTTTGC  2649
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  TGAAGAGGAAACTCTGTCACTCCTAGCTCATACACCCATTCTCTTTGC  2520

Qy  2650  TTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT  2709
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT  2580

Qy  2710  TATCTTCTCCCTCCCCTCCCCTTCTTCCATCCTCCATCTCTTTGAATTCCTCAT  2769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TATCTTCTCCCTCCCCTCCCCTTCTTCCATCCTCCATCTCTTTGAATTCCTCAT  2640

Qy  2770  CCCTCCATCTCAATCCCGTATCTACGCACCCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA  2829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  CCCTCCATCTCAATCCCGTATCTACGCACCCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA  2700

Qy  2830  GTATCACATCACACAAAAAGGAACAAAAGCGAAACACACAAAAACCAGCCTCAACTTACACTT  2889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  GTATCACATCACACAAAAAGGAACAAAAGCGAAACACACAAAAACCAGCCTCAACTTACACTT  2760

Qy  2890  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTACGGTTTTGGAAGAGAGAAAACA  2949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGAAAACA  2820

Qy  2950  GGAACCCACCAAAACCAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGT  3009
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  GGAACCCACCAAAACCAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGAATGT  2880

Qy  3010  ATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGCAAAATGATTCCTTTC  3069
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  ATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGCAAAATGATTCCTTTC  2940

Qy  3070  TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCCAGGGGCGTT  3129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  TTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCCAGGGGCGTT  3000

Qy  3130  AAATTCACAGATTTTTTAACGAGAAAAACACAGAAGAAGCTACCTCAGGTGTTTTTA  3189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  AAATTCACAGATTTTTTAACGAGAAAAACACAGAAGAAGCTACCTCAGGTGTTTTTA  3060

Qy  3190  CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCAT  3249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  CCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCAT  3120

Qy  3250  TTTTTAATTTTTTAATAAAAATGAGTTGGAAAAATAAGATATCAACTGCCAGCCTG  3309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  TTTTTAATTTTTTAATAAAAATGAGTTGGAAAAATAAGATATCAACTGCCAGCCTG  3180

Qy  3310  GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGCCTAGCCAAGAA  3369
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  GAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGCCTAGCCAAGAA  3240

Qy  3370  CCNATATGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
          ||||||||||||||||||||||||||||||||||||||||||
Db  3241  CCNATATGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3283
```

RESULT 4
US-09-814-353-20241
; Sequence 20241, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James

```
Qy  1870  GGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAGAA  1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2107  GGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCTGACAGAA  2166

Qy  1930  TGAGACCAAACGCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAGGAATGAGAA  1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2167  TGAGACCAAACGCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAGGAATGAGAA  2226

Qy  1990  GTCTGCGGAGGCGGCGGCCAGGGACTCTGCCGAGGGCCCTGAGAACCCAGGGGCCGAGGAGGGG  2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2227  GTCTGCGGAGGCGGCGGCCAGGGACTCTGCCGAGGGCCCTGAGAACCCAGGGGCCGAGGAGGGG  2286

Qy  2050  GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCTCCCGCCAGGGC  2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2287  GCGGGGAAGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGCCTCCCGCCTCCCGCCAGGGC  2346

Qy  2110  TTCTGCAGGCTTCAGCCATCCACTTCACCATCCTCGATCTCTCCTGAACTCCCACGA  2169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2347  TTCTGCAGGCTTCAGCCATCCACTTCACCATCCTCGATCTCTCCTGAACTCCCACGA  2406

Qy  2170  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2407  CGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2466

Qy  2230  ACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTATACATATTAGAAAGGAAGATG  2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2467  ACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTATACATATTAGAAAGGAAGATG  2526

Qy  2290  TTAAGATATGTGGCCTGTGTGGGTTACACAGGGGTGCCTGCAGCGGTAATATATTTTAGAAAT  2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2527  TTAAGATATGTGGCCTGTGTGGGTTACACAGGGGTGCCTGCAGCGGTAATATATTTTAGAAAT  2586

Qy  2350  AATATATCAAATAACTCAACTAACTAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTTCTTT  2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2587  AATATATCAAATAACTCAACTAACTAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTTCTTT  2646

Qy  2410  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTTGGGAGGTCTCACGG  2469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2647  TTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTTGGGAGGTCTCACGG  2706

Qy  2470  TGTGAGAGAGGAGCTTTGAGGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2707  TGTGAGAGAGGAGCTTTGAGGGCCACCCGACAAAATTCACCCAGAGGGAAATCTCGTCGGA  2766

Qy  2530  AGGACACTCACGGCAGTTCTCGATCACCTGTATGTCAACAGAGAGGGATACCGTCTCCT  2589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2767  AGGACACTCACGGCAGTTCTCGATCACCTGTATGTCAACAGAGAGGGATACCGTCTCCT  2826

Qy  2590  TGAAGAGGAAACTCTGTCACTCCTGCTCAGCTCATACACCCATTTCTCTTGC  2649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2827  TGAAGAGGAAACTCTGTCACTCCTGCTCAGCTCATACACCCATTTCTCTTGC  2886

Qy  2650  TTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT  2709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2887  TTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTCTGTT  2946

Qy  2710  TATCTCTCCCCTCCCCTCCCCCCTTCTTCCTCATCTCCATTCTTTTGAATTCCTCAT  2769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2947  TATCTCTCCCCTCCCCTCCCCCCTTCTTCCTCATCTCCATTCTTTTGAATTCCTCAT  3006

Qy  2770  CCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCCCCCAGGCAAAGCAGTAGCTCTGA  2829
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  3007  CCCTCCATCTCAATCCCGTATCTACGGCA--CCCCCCCCCCCCAGGCAAAGCAGTAGCTCTGA  3065

Qy  2830  GTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAACCAGGCTCAACTTACACTT  2889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3066  GTATCACATCACACAAAAGGAACAAAAGCGAAACACACAAACCAGGCTCAACTTACACTT  3125

Qy  2890  GGTTACTCAAAGAACAAGAGTCAATGGACGTTTTGGAGGAGAGAAAACA  2949
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3126  GGTTACTCAAAGAACAAGAGTCAATGGACGTTTTGGAGGAGAGAAAACA  3185
```

```
Qy  2950  GGAACCCACCAAACCAACCAACCAATCAACCAAACAAACAAGAAGAAGAAAATTCCAACAATGAAAGAATGT  3009
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3186  GGAACCCACCAAACCAACCAACCAATCAACCAAACAAACAAGAAGAAGAAAATTCCAACAATGAAAGAATGT  3245

Qy  3010  ATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC  3069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3246  ATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC  3305

Qy  3070  TTT-AAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGT  3128
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3306  TTTAAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGGCGT  3365

Qy  3129  TAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGAGCTACCTCAGTGTTTTT  3188
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3366  TAAATTCACAGATTTTTTAAGAGAGAAAAAAACACAGAGAAGAGCTACCTCAGTGTTTTT  3425

Qy  3189  ACCTCAGCACCTTGCTCTTGTCTTGTGTTTCCTTAGAGATTTTGTAAAGTTGATAGTTGGAGCA  3248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3426  ACCTCAGCACCTTGCTCTTGTCTTGTGTTTCCTTAGAGATTTTGTAAAGTTGATAGTTGGACA  3485

Qy  3249  TTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAGAATATCAACTGCCAGCCT  3308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3486  TTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAGAATATCAACTGCCAGCCT  3545

Qy  3309  GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTCGGACTTGCTTCCGCTAGCCAAGA  3368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3546  GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTCGAATTGCTTCCGCTAGCCAAGA  3605

Qy  3369  ACC  3371
          |||
Db  3606  ACC  3608
```

RESULT 5
US-10-097-340-146
; Sequence 146, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Marijula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19

```
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 146
; LENGTH: 3667
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-097-340-146

Query Match         53.0%;  Score 1810;  DB 15;  Length 3667;
Best Local Similarity  99.9%;  Pred. No. 0;
Matches  2100;  Conservative  0;  Mismatches  1;  Indels  2;  Gaps  2;

Qy   1270  ACCCACTCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT  1329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507  ACCCACTCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCAT  1566

Qy   1330  CATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGC  1389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1567  CATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGC  1626

Qy   1390  GCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCCTCT  1449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1627  GCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCCTCT  1686

Qy   1450  ATCAAGAGTTGCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG  1509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1687  ATCAAGAGTTGCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGG  1746

Qy   1510  CCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAAAC  1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1747  CCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAAAC  1806

Qy   1570  TTCTTTAACCCCAAAGAAGAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCACA  1629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1807  TTCTTTAACCCCAAAGAAGAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCACA  1866

Qy   1630  GCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT  1689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1867  GCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGT  1926

Qy   1690  GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAGGAGTGATGTCGTCAGA  1749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1927  GCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAGGAGTGATGTCGTCAGA  1986

Qy   1750  ATTATCGGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA  1809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1987  ATTATCGGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAA  2046

Qy   1810  CAGGTGAAGCAGCAGGAGCGAGCAGAAATACCCTGCCTCACAGCGCCAGCAAGTGA  1869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2047  CAGGTGAAGCAGCAGGAGCGAGCAGAAATACCCTGCCTCACAGCGCCAGCAAGTGA  2105

Qy   1870  GGCTCCCACCAGGCACCAGGCGACGCAGCCAGATCGGGAGCCAAACCATCTGAGGAATGTG  1929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2107  GGCTCCCACCAGGCACCAGGCGACGCAGCCAGATCGGGAGCCAAACCATCTGAGGAATGTG  2166

Qy   1930  TGAGACCAAACGCAGCCAGCCAGCCGGGAGCCAGCAAACCAAAGACCATCTGAGGAATGAA  1989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2167  TGAGACCAAACGCAGCCAGCCAGCCGGGAGCCAGCAAACCAAAGACCATCTGAGGAATGAA  2226

Qy   1990  GTCTGCGGAGGTCAGGTTGCCAGGACTCTGCCGAGGGCCTGAGGGGGCCGAGGAGGG  2049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2227  GTCTGCGGAGGTCAGGTTGCCAGGACTCTGCCGAGGGCCTGAGGGGGCCGAGGAGGG  2286

Qy   2050  GCGGGGAAGGTCAGCCAGGTCAGCCATCCAACCACCACCGAGCCCCGCCTCCCGAACTCCCACGA  2109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2287  GCGGGGAAGGTCAGCCAGGTCAGCCATCCAACCACCACCGAGCCCCGCCTCCCGAACTCCCACGA  2346

Qy   2110  TTCTGCAGGCTTCAGCGACTTCACCATCCATCGGATCTCTCGGAACTCCCACGA  2169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2347  TTCTGCAGGCTTCAGCGACTTCACCATCCATCGGATCTCTCGGAACTCCCACGA  2406

Qy   2170  CGCTATCCCTTTTAGTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2407  CGCTATCCCTTTTAGTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCAC  2466

Qy   2230  ACCCTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2467  ACCCTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATG  2526

Qy   2290  TTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCCAGCGGGTAATATATTTTAGAAAT  2349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2527  TTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCCAGCGGGTAATATATTTTAGAAAT  2586

Qy   2350  AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTCTTT  2409
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2587  AATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTCTTT  2646

Qy   2410  TTAAAGAGAAAGCAGGCTTTTCTAGACTTAAAGATAAAGTCTTTGGGAGGTCTCACGG  2469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2647  TTAAAGAGAAAGCAGGCTTTTCTAGACTTAAAGATAAAGTCTTTGGGAGGTCTCACGG  2706

Qy   2470  TGTAGAGGAGCTTTGAGGGCCACCCGACAAAATCACCCAGAGGAGAAATCTCGTCGGA  2529
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2707  TGTAGAGGAGCTTTGAGGGCCACCCGACAAAATCACCCAGAGGAGAAATCTCGTCGGA  2766

Qy   2530  AGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAGGGATACCGTCGTCCT  2589
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2767  AGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAGGGATACCGTCGTCCT  2826

Qy   2590  TGAAGAGGAAACTTCTGTCACTCCTGCCTGCTCAGCTCATCACCCATTTCTCTTTTTGC  2649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2827  TGAAGAGGAAACTTCTGTCACTCCTGCCTGCTCAGCTCATCACCCATTTCTCTTTTTGC  2886

Qy   2650  TTCACAGGTTTAAACTGGTTTTGCATATGCTATATATAATTCTCTGTCTCTCTCGTT  2709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2887  TTCACAGGTTTAAACTGGTTTTGCATATGCTATATATAATTCTCTGTCTCTCTCGTT  2946

Qy   2710  TATCTCCATCTCAATCCCGTATCTACGGACCCCCCCCCCCCCCCCCCATTCCTCAT  2769
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   2947  TATCTCCATCTCAATCCCGTATCTACGGCCA--CCCCCCCCCCCCCCCCATTCCTCAT  3006

Qy   2770  CCCTCCATCTCAATCCCGTATCGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTT  2829
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3007  CCCTCCATCTCAATCCCGTATCGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTT  3065

Qy   2830  GTATCACATCACACAAAAGCGAAACACAAAAGGAACAAAAGGCAAAATTGTACAA  2889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066  GTATCACATCACACAAAAGCGAAACACAAAAGGAACAAAAGGCAAAATTGTACAA  3125

Qy   2890  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAGAGAGAAAACA  2949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3126  GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAGAGAGAAAACA  3185

Qy   2950  GGAACCCACCAAAACCAACCAATCAACCAAAACAAACAAGAAAAATTCCACAATGAAAGAATGT  3009
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3186  GGAACCCACCAAAACCAACCAATCAACCAAAACAAACAAGAAAAATTCCACAATGAAAGAATGT  3245

Qy   3010  ATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAGAAATGATTCCTTTC  3069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3246  ATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAGAAATGATTCCTTTC  3305

Qy   3070  TTT-AAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGCCCAGGGGCGT  3128
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3306  TTTAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGCCCAGGGGCGT  3365

Qy   3129  TAAATTCACAGATTTTTTCCCTTTAACGAGAAAAACACACAGAGAAGCAAAGCTTAAGCTCCAGGCGCT  3188
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3366  TAAATTCACAGATTTTTTCCCTTTAACGAGAAAAACACACAGAGAAGCAAAGCTTAAGCTCCAGGCGCT  3425

Qy   3189  ACCTCAGCCACTTCAGATTTGTTCTCTGTTTCCCTTTGAGATTTTGTAAAGCTGATAGTTGGAGCA  3248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3426  ACCTCAGCCACTTCAGATTTGTTCTCTGTTTCCCTTTGAGATTTTGTAAAGCTGATAGTTGGAGCA  3485

Qy   3249  TTTTTTTTATTTTTTAATAAAAATGAGTTGAAAAAAATAAGATATCAACTGCCAGCCT  3308
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 6
US-10-648-593-48
; Sequence 48, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
;   TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
;   TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
;   FILE REFERENCE: D0273 NP
;   CURRENT APPLICATION NUMBER: US/10/648,593
;   CURRENT FILING DATE: 2003-08-26
;   PRIOR APPLICATION NUMBER: 60/406,385
;   PRIOR FILING DATE: 2002-08-27
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 48
;   LENGTH: 3667
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-648-593-48

Query Match   53.0%; Score 1810; DB 17; Length 3667;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2100; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1270 ACCACTCCGATACTTCTCCAGCTCGTACCCCATCACCCGTTGGCCCGTTCCGCAT 1329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1507 ACCCACTCCGATACTTCTCCAGCTCGTACCCCATCACCCGTTGGCCCGTTCCGCAT 1566

Qy 1330 CATCACTCTTATCCAGAGCAGGAGAGTTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGC 1389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1567 CATCACTCTTATCCAGAGCAGGAGAGTTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGC 1626

Qy 1390 GCCATCATCGGGAAGAAGAAGGGGGCACACATCAAAACAGCTCGCGGAGATTCGCCGGAGCCTCT 1449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1627 GCCATCATCGGGAAGAAGAAGGGGGCACACATCAAAACAGCTCGCGGAGATTCGCCGGAGCCTCT 1686

Qy 1450 ATCAAGATTGCCCCTGCGGAAGGCCAGACGTCAGCGAAGGATGGTCATCATCACCGGG 1509
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1687 ATCAAGATTGCCCCTGCGGAAGGCCAGACGTCAGCGAAGGATGGTCATCATCACCGGG 1746

Qy 1510 CCACCCGGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGAAACTGAAAGAGAGAAAC 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1747 CCACCCGGAAGCCCAGTTCAAGGCCCAGGGACGATCTTTGGAAACTGAAAGAGAGAAAC 1806

Qy 1570 TTCTTTAACCCCAAAGAAGAAGAGTGAAGCTGAAGCTGAAGGCCATATCAGAGTGCCCTCTTCCACA 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1807 TTCTTTAACCCCAAAGAAGAAGAGTGAAGCTGAAGCTGAAGGCCATATCAGAGTGCCCTCTTCCACA 1866

Qy 1630 GCTGGCCGGTGATTGGCAAAGGTGGCAAGACCGTTGAACGAACTGCAGAACTTAACCAGT 1689
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1867 GCTGGGCCGGTGATTGGCAAAGGTGGCAAGACCGTTGAACGAACTGCAGAACTTAACCAGT 1926

Qy 1690 GCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAAAATGAGGAAGTCGATCGTCAGA 1749
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1927 GCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAAAATGAGGAAGTCGATCGTCAGA 1986

Qy 1750 ATTATCGGGCACTTCTTTGCTAGCCAGACTCGCACAGCGCAAGATCAGGGAAATTGTACAA 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1987 ATTATCGGGCACTTCTTTGCTAGCCAGACTCGCACAGCGCAAGATCAGGGAAATTGTACAA 2046

Db 3486 TTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCT 3545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3309 GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTCTTCCGCTAGCCAAGA 3368

Db 3546 GGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTCTTCCGCTAGCCAAGA 3605
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3369 ACC 3371

Db 3606 ACC 3608

Db 1810 CAGGTGAAGCAGCAGGAGCAGGAGCAGCAGAAATACCCTCAGGGAGTCGCTCGACAGCGCGAGCAAGTGA 1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2047 CAGGTGAAGCAGCAGGAGCAGGAGCAGCAGAAATACCCTCAGGGAGTCGCTCACAGCGCGAGCAAGTGA 2106

Db 1870 GGCTCCACAGCACCAGGCACACCAGGGACTCGCCGAGCCCTTCCAACACTGACAGAA 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2107 GGCTCCACAGCACCAGGCACACCAGGGACTCGCCGAGCCCTTCCAACACTGACAGAA 2166

Db 1930 TGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2167 TGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAA 2226

Db 1990 GTCTGCGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGGAGGG 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2227 GTCTGCGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGAGGAGGG 2286

Db 2050 GCGGGGAGGTCAGCCAGGTTTGCCAGACACCAGGGCCCGCTCCTCCTCGAACTCCCACGA 2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2287 GCGGGGAGGTCAGCCAGGTTTGCCAGACCACCACCGAGCCCCGCTCCTCCTCGAACTCCCACGGC 2346

Db 2110 TTCTGCAGGCTTCAGCAATCCACCATCCACCACTGGATCTCTCCTGAACTCCCACGA 2169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2347 TTCTGCAGGCTTCAGCCATCCACCATCCACCACTGGATCTCTCCTGAACTCCCACGA 2406

Db 2170 CGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAAATGCAC 2229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2407 CGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAAATGCAC 2466

Db 2230 ACCCTTTTCTGTGTGCAAATCGTCTCGTGGGTTGGGTGCCTGCAGCCGGTAATATATTTTAGAAAT 2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2467 ACCCTTTTCTGTGTGCAAATCGTCTCGTGGGTTGGGTGCCTGCAGCCGGTAATATATTTTAGAAAT 2526

Db 2290 TTAAGATATGTGGCCTGTGGGTTGGGTTACACAGGGTGCCTGCAGCGGTGCCTGGTAATATTTTTTTTTCTT 2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2527 TTAAGATATGTGGCCTGTGGGTTGGGTTACACAGGGTGCTGCAGCGGTGCCTGGTAATATTTTTAGAAAT 2586

Db 2350 AATATATCAAATAACTAACTAACTAACTCAATTTTTAATCAATTATTAATTATTAATTTTTTTTTCTT 2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2587 AATATATCAAATAACTAACTAACTAACTCAATTTTTAATCAATTATTAATTAATTTTTTTTTCTT 2646

Db 2410 TTAAAGAAAGCAGAGCGTTTTCTAGACTTTAAAGAAGAATAAAGTCTTTGGGAGGTCTCACGG 2469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2647 TTAAAGAAAGCAGAGCGTTTTCTAGACTTTAAAGAAGAATAAAGTCTTTGGGAGGTCTCACGG 2706

Db 2470 TGTAGAGAGGAGTTTGAGCGCACCCGCACCCCGACCAAAATTCACCGGAGGGAAATCTCGTCGGA 2529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2707 TGTAGAGAGGAGTTTGAGCGCACCCGCACCCCGACCAAAATTCACCCGAGGGAAATCTCGTCGGA 2766

Db 2530 AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGCAAGGGGTATACCGTCTCCT 2589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2767 AGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGGAGGGATATACCGTCTCCT 2826

Db 2590 TGAGAGGAAACTTCTGTCACTCGTCTCAGCTCATCACACCCATTCTCTCTTTGC 2649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2827 TGAGAGGAAACTTCTGTCACTCGTCTCATGCCGTGTCTGTCACACCCATTCTCTCTTTGC 2886

Db 2650 TTCACAGGTTTAAACTGGTTTTTTTGCATACTGGTTTTTTGCATACTGTCTGTCTCTCTGTT 2709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2887 TTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTTTTTTGCATACTGTCTGTCTCTCTGTT 2946

Db 2710 TATCTCTCCCCTCCCCCCTTCTTCTCCATCTCCATCTCTTTGAATTCCTCAT 2769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2947 TATCTCTCCCCTCCCCCCTTCTTCTCCATCTCCATCTCTTTGAATTCCTCAT 3006

Db 2770 CCCTCCATCTCAATCCGTATCTCACGCACCCCCCCCCCCCCCCCCCAGGCAAAGCAGTGCTCTGA 2829
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Qy 3007 CCCTCCATCTCAATCCGTATCTCACGCACCCCCCCCCCCCCCCC-CCCCCAGGCAAAGCAGTGCTCGA 3065

Db 2830 GTATCACATCACACAAAAGGCGAAACACAAAACCAGCCTCAACTTACACTT 2889
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3066 GTATCACATCACACAAAAGGCGAAACACAAAACCAGCCTCAACTTACACTT 3125

Db 2890 GGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGAAGAGAGAAAACA 2949

```
Db  216  AACTTCTTTAACCCAAAGAAGAAGAGTGAAGCGCCATATCAGAGTGCCTCTCC  275
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3126  GGTTACTCAAAGAACAAGACCAACCAATCAATGGTACTGTCCTAGCGTTTGGAAGAGGAAAACA 3185

Qy 1627  ACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGCAGAACTTAACC 1686
         ||||||||||||||||||||||||||||||||||||||||||
Db  276  GGAACCCACCAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAAATGT 335
Qy 2950  GGAACCCACCAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAAATGT 3009

Qy 1687  AGTGCAGAAGCTCATCGTGCCTCGTGACCAAACGGCCAGAGAAATGAGGAGAGTGATCGTC 1746
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  336  AGTGCAGAAGCTCATCGTGCCTCGTGACCAAACGCCAGATGAAATGAGGAGAGTGATCGTC 395
Db 3186  GGAACCCACCAAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATGAAAGAAATGT 3245

Qy 1747  AGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCGCAAGATCAGGGAAAATTGTA 1806
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  396  AGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCGCAAGATCAGGGAAAATTGTA 455
Qy 3010  ATTTTGTCTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3069

Db  456  CAACAGTGAAGCAGCAGCAGGAGCAGAGAATACCCTCAGGGAGTGCCTCACAGCGCAGCAAG 515
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 1807  CAACAGTGAAGCAGCAGCAGGAGCAGAGAATACCCTCAGGGAGTGCCTCGCTCACAGCGCAGCAAG 1866
Db 3246  ATTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTC 3305

Qy 1867  TGAGGCTCCCACAGGACCACCAGCAAACAACGGACGATGAATGTAGCCCTTCCAACACCTGACA 1926
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  516  TGAGGCTCCCACAGGACCACCAGCAACGGACGATGAATGTAGCCCTTCCAACACCTGACA 575
Qy 3070  TTT-AAAAAAAAAATGTGGAGGAAGTAGAAAATTACCAAGGTTGTTGGCCCCAGGGGGT 3128

Db  576  GAATGAGACCAAACGCAGCCAGATCGGGAGCAAACAAAGACCATCTGAGGAATGA 635
         |||||||||||||||||||||||||||||||||||||
Qy 1927  GAATGAGACCAAACGCAGCCAGATCGGGAGCAAACAAAGACCATCTGAGGAATGA 1986
Db 3306  TTTAAAAAAAAAATGTGGAGGAAGTAGAAAATTACCAAGGTTGTTGGCCCCAGGGGCT 3365

Qy 1987  GAAGTCTGCGGAGGCGGCCAGGAGACTCTGCCAGGACCACCCGAGAACCACCAGGGCCGAGGA 2046
         |||||||||||||||||||||||||||||||||||||||||||||
Db  636  GAAGTCTGCGGAGGCGGCCAGGAGACTCTGCCGAGGGCCCTGAGAACCACCAGGGCCGAGGA 695
Qy 3129  TAAATTCACAGATTTTTTAACGAGAAAAAACACACAGAGAGAAGAGCTACCTTCAGGTGTTTTT 3188

Db  696  GGGGCGGGAGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCATCCACTCCTGAACTCCCA 755
         |||||||||||||||||||||||||||||||||||||||
Qy 2047  GGGGCGGGAGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCATCCACTCCTGAACTCCCA 2106
Db 3366  TAAATTCACAGATTTTTTTTAACGAGAAAAAACACACAGAGAGAAGAGCTACCTTCAGGTGTTTTT 3425

Qy 2107  GGCTTCTGCAGGCTTCAGCCATCCACTTCACCTCACTCGGATCTCTCCTGAACTCCCA 2166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  756  GGCTTCTGCAGGCTTCAGCCATCCACTTCACCTCACTCGGATCTCTCCTGAACTCCCA 815
Qy 3189  ACCTCAGCACCTTGCTCTTGTTCCCTTAGAGATTTTGTAAAGCTAAAGTTGTGATAGTTGGAGCA 3248

Db  816  CGAACGCTATCCTTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAATG 875
         |||||||||||||||||||||||||||||||||||||||||||
Qy 2167  CGAACGCTATCCTTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCCAAGCAAAATG 2226
Db 3426  ACCTCAGCACCTTGATCTCTTTGTTCCCTTAGAGATTTTGTAAAGCTAAAGTTGTGATAGTTGGAGCA 3485

Qy 2227  CACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAGGGGAAG 2286
         |||||||||||||||||||||||||||||||||||||||||||
Db  876  CACACCCTTTTTCTGTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAG 935
Qy 3249  TTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAAATAAGATATCAACTGCCAGCCT 3308

Db  936  ATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATTTTAGA 995
         ||||||||||||||||||||||||||||||||||||||||||
Qy 2287  ATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATTTTAGA 2346
Db 3486  TTTTTTTATTTTTTTTATAAAAATGAGTTGGAAAAAAAAAATAAGATAATCAACTGCCAGCCT 3545

Qy 2347  AATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAAATTTTTTTTTC 2406
         |||||||||||||||||||||||||||||||||||||||
Db  996  AATAATATATCAAATAACTCAACTAACTCCAATTTAATCAATTATTAAATTTTTTTTTC 1055
Qy 3309  GGAGAAGGGTGACAGTCCAAGTGTTCAAACAGCTGTTCTTCCGCTAGCCAAGA 3368

Db 1056  TTTTTAAAGAGAAAGCAGAGCTTTTGAGGACTTAAGGTTTTTGAGGAGGTCTCA 1115
         ||||||||||||||||||||||||||||||||||||||||
Qy 2407  TTTTTAAAGAGAAAGCAGAGCTTTTGAGGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCA 2466
Db 3546  GGAGAAGGTGACAGTCGACAGTCCAAGTGTTCAAACAGCTGTTCTTCCGCTAGCCAAGA 3605

Qy 2467  CGGTGTGAGAGAGGAGCTTTGAGGCCACCCCACAAAATTCACCCGAGGGAAAATCCGTC 2526
         |||||||||||||||||||||||||||||||||||||||||||||
Db 1116  CGGTGTTAGAGAGGAGCTTTGAGGCCACCCCACAAAATTCACCCGAGGGAAATCCGTC 1175
Qy 1387  GGGCCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCCGAGATTCGCCGGAGCC 1446

Db 1176  GGAAGGACACTCACCGGCAGTTCTGGATCACCGGTCATCACC 1235
         |||||||||||||||||||||||||||||||||||||||
Qy 2527  GGAAGGACACTCACCGGCAGTTCTGGATCACCGGTCATCATCACC 2586
Db   36  GGGCGGCCCATCATCGGGAAGAAGGGGGCGCACACATCAAACAGCTGGCCGAGATTCGCCGGAGCC 95

Qy 2587  CCTTGAAGAGGAAACTTCTGTCACTTCCTCATGCCTTGGCTCAGCTGCCTCATCACCCATTTCTCTT 2646
         |||||||||||||||||||||||||||||||||||||||||||
Db 1236  CCTTGAAGAGGAAACTTCTGTCACTTCCTCATGCCTTGGCTCAGCTGCCTCATCACCCATTTCTCTT 1295
Qy 1447  TCTATCAAGATTGCCCCTGCGGAAGGGCCCAGACGTCAGCAGGGACCCAGGACCCAGGATGGTCATCATCACC 1506

Db 1296  TGCTTCACAGGTTTTAAACTGGTTTTTTGCATACGTCGTATATAATTCTCTGTCTCTCT 1355
         |||||||||||||||||||||||||||||||||||||||||||
Qy 2647  TGCTTCACAGGTTTTAAACTGGTTTTTTGCATACGTCGTATATAATTCTCTGTCTCTCT 2706
Db  156  GGGCGGCACCGGAAGGCCAGTTCAAGGGCCCAGGACGTCAGCAGGGACCCAGGACCCAGGATGGTC 215
```

```
Qy 3369  ACC 3371
         |||
Db 3606  ACC 3608
```

RESULT 7
US-09-764-864-332

```
; Sequence 332, Application US/09764864
; Patent No. US20020137753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
;   LENGTH: 2290
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-332
```

Query Match   43.8%;   Score 1496;   DB 9;   Length 2290;
Best Local Similarity   99.7%;   Pred. No. 0;
Matches 1966;   Conservative   0;   Mismatches   1;   Indels   4;   Gaps   3;

US-09-764-864-749

```
Query Match             11.8%;   Score 401;  DB 9;  Length 1186;
Best Local Similarity   99.4%;   Pred. No. 3.2e-194;
Matches   671;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;
```

```
Qy   1864  AAGTGAGGCTCCCACAGGCACCAGCAAAACACGGATGAATGTAGCCTTCCAACACCTG  1923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    512  AAGTGAGGCTCCCACAGGCACCAGCAAAACACGGATGAATGTAGCCCTTCCAACACCTG  571

Qy   1924  ACAGAATGAGACCAAACGCAGCCAGCCAGGATCGGGAGCAAACCAAAGACCATCTGAGGAA  1983
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    572  ACAGAATGAGACCAAACGCAGCCAGCCAGGATCGGGAGCAAACCAAAGACCATCTGAGGAA  631

Qy   1984  TGAGAAGTCTGCGAGGCGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGA  2043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    632  TGAGAAGTCTGCGAGGCGGCGGCCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGA  691

Qy   2044  GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGGCTCCCGCCCCCC  2103
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    692  GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCCGGCTCCCGCCCCCC  751

Qy   2104  CAGGGCTTCTGCAGGCTCAGCCAT-CACTTCACCATCCACTCGGATCTCTCCTGAACT  2162
           ||||||||||||||||||||||   || |||||||||||||||||||||||||||||
Db    752  CAGGGCTTCTGCAGGCTTCAGCGCATGTAGTTGAACTAACATAGTGAACGTGTTCAAGCCAAGCAA  811

Qy   2163  CCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAGCCAAGCAA  2222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    812  CCCACGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAGCCAAGCAA  871

Qy   2223  AATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGG  2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    872  AATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGG  931

Qy   2283  GAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCGAGCGGGTAATATATT  2342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    932  GAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCGAGCGGGTAATATATT  991

Qy   2343  TAGAAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATATTATTAATTTTTT  2402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992  TAGAAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATATTATTAATTTTTT  1051

Qy   2403  TTTCTTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT  2462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1052  TTTCTTTTTTAAAGAGAAAGCMGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGT  1111

Qy   2463  CTCACGGTGTAGAGAGGACACT  2522
           ||||||||||||||||||||||
Db   1112  CTCACGGTGTAGAGAGGACACT  1171

Qy   2523  CGTCGGAAGGACACT  2537
           |||||||||||||||
Db   1172  CGTCGGAAGGACACT  1186
```

RESULT 9

US-09-814-353-14725

```
; Sequence 14725, Application US/09814353
; Publication No. US2003016581A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT: Thompson, Pamela
;  APPLICANT: Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
```

US-09-764-864-749

```
Qy   2707  GTTTATCTCTCCCCTCCCTTCTTCTTCCATCTCCATTCTTTGAATTCCT  2766
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1356  G-TATCTCTCCCCCTCCCTCCCCTTCTTCTCCATCTCCATTCTTTGAATTCCT  1414

Qy   2767  CATCCTCCATCTCAATCCGTATCTACGCA--CCCCCCCCCCCCCCCAGGCAAAGCAGTGC  2824
           ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db   1415  CATCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCCCCCCCAGGCAAAGCAGTGC  1474

Qy   2825  TCTGAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAACCAGCCTCAACTTA  2884
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475  TCTGAGTATCACATCACACAAAAGGAACAAAAGGCAAACACACAACCAGCCTCAACTTA  1534

Qy   2885  CACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGTGGAAGAGGA  2944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535  CACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGTGGAAGAGGA  1594

Qy   2945  AAACAGGAACCCACCAACCAAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAG  3004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595  AAACAGGAACCCACCAACCAAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAG  1654

Qy   3005  AATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTC  3064
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655  AATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTC  1714

Qy   3065  CTTTCTTT-AAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGGCCCAG  3123
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1715  CTTTCTTTAAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGTTGTTGGCCCAG  1774

Qy   3124  GGCGTTAAATTCACAGATTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGGTG  3183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775  GGCGTTAAATTCACAGATTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGGTG  1834

Qy   3184  TTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTG  3243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835  TTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTG  1894

Qy   3244  GAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCC  3303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895  GAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCC  1954

Qy   3304  AGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGCT  3354
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955  AGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCT  2005
```

RESULT 8

US-09-764-864-749

```
; Sequence 749, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 749
;  LENGTH: 1186
;  TYPE: DNA
;  ORGANISM: Homo sapiens
FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (511)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: SITE
;  LOCATION: (774)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: SITE
;  LOCATION: (777)
;  OTHER INFORMATION: n equals a,t,g, or c
```

; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14725
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-14725

Query Match   11.6%;   Score 397; DB 10; Length 582;
Best Local Similarity 99.8%; Pred. No. 3.5e-192;
Matches 517; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 2855 AAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAA 2914
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 AAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAA   80

Qy 2915 TGGTACTTGTTCCTAGCCGTTTGGAAGAGGAAAACAGGAAACCCACCAACCAACCAATCAA 2974
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 TGGTACTTGTTCCTAGCCGTTTGGAAGAGGAAAACAGGAAACCCACCAACCAACCAATCAA   140

Qy 2975 CCAAACAAAGAAAAAATTCCACAATGAAGAATGAAAGAATGTATTTTGTCTTTTTGGTGT 3034
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 CCAAACAAAGAAAAAATTCCACAATGAAGAATGAAAGAATGTATTTTGTCTTTTTGGTGT   200

Qy 3035 ATAAGCCATCAATATTCAGCAGAAAATGATTCCTTCTTT-AAAAAAAAAAAATGTGGAGAA 3093
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   201 ATAAGCCATCAATATTCAGCAGAAAATGATTCCTTCTTTCTTTAAAAAAAAAAAATGTGGAGAA   260

Qy 3094 AGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAG 3153
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 AGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAG   320

Qy 3154 AAAAACACACAGAGAAGCTACCTCAGGTGTTTTACCCTGACACCTTGCTCTTGTGTTT 3213
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 AAAAACACACAGAGAAGCTACCTCAGGTGTTTTACCCTGACACCTTGCTCTTGTGTTT   380

Qy 3214 CCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTATTTTTTTAATAAAAATG 3273
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 CCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTATTTTTTTAATAAAAATG   440

Qy 3274 AGTTGGAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGC 3333
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441 AGTTGGAAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGC   500

Qy 3334 AACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC 3371
          ||||||||||||||||||||||||||||||||||||||
Db   501 AACAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC   538

RESULT 10
US-09-764-864-330
; Sequence 330, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 330
; LENGTH: 822
; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (441)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (455)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-330

Query Match   11.4%;   Score 389; DB 9; Length 822;
Best Local Similarity 99.8%; Pred. No. 4.5e-188;
Matches 439; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy   779 CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCGAAGCATGCCGCA   838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCGAAGCATGCCGCA   60

Qy   839 TGATTCTTGAAATCATCGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC   898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGATTCTTGAAATCATCGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC   120

Qy   899 TGAAAATCTTGGCACACAAATGCTTGGTTGGAAGACTGATTGGAAAAAGAGAGGCAGAAATT   958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGAAAATCTTGGCACACAAATGCTTGGTTGGAAGACTGATTGGAAAAAGAGAGGCAGAAATT   180

Qy   959 TGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGA 1018
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   181 TGAAGAAAATTGAACATGAAACAGGGACCAAGRTAACAATCTCATCTTTGCAGGATTTGA   240

Qy 1019 GCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTCGCCAGTG 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTCGCCAGTG   300

Qy 1079 CTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTA 1138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTA   360

Qy 1139 ACCAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCACTTGCATCTTTCAACAG 1198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ACCAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCACTTGCATCTTTCAACAG   420

Qy 1199 GACTGTTCCGTGCTATCTCCA 1218
          |||||||||||||||||||||
Db   421 GACTGTTCCGTGCTATCTCCA   440

RESULT 11
US-09-814-353-1995
; Sequence 1995, Application US/0981435353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1995
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-1995

Query Match   10.2%;   Score 347;   DB 10;   Length 579;
Best Local Similarity   99.8%;   Pred. No. 1.5e-166;
Matches 467;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

Qy 2846 AAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 2905
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8 AAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 67

Qy 2906 AAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 2965
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 AAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGGAAAACAGGAACCCACCAAACCA 127

Qy 2966 ACCAATCAACCAAACAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 3025
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 ACCAATCAACCAAACAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 187

Qy 3026 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTT-AAAAAAAAAAT 3084
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   188 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAAAAAAT 247

Qy 3085 GTGGAGGAAAGTAGAAATTTACCAAGGTTGTGGCCCAGGGCGTTAAATTCACAGATTTT 3144
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 GTGGAGGAAAGTAGAAATTTACCAAGGTTGTGGCCCAGGGCGTTAAATTCACAGATTTT 307

Qy 3145 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 3204
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 367

Qy 3205 CTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTTA 3264
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   368 CTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTTA 427

Qy 3265 ATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCTGGAG 3312
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   428 ATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCTGGAG 475

RESULT 12
US-09-814-353-8341
; Sequence 8341, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8341
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-8341

Query Match   10.2%;   Score 347;   DB 10;   Length 579;
Best Local Similarity   99.8%;   Pred. No. 1.5e-166;
Matches 467;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

Qy 2846 AAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 2905
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8 AAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAAC 67

Qy 2906 AAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAAAACAGGAACCCACCAAACCA 2965
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 AAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGAGGAAAACAGGAACCCACCAAACCA 127

Qy 2966 ACCAATCAACCAAACAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 3025
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 ACCAATCAACCAAACAAACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTTGCA 187

Qy 3026 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTT-AAAAAAAAAAT 3084
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   188 TTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAAAAAAT 247

Qy 3085 GTGGAGGAAAGTAGAAATTTACCAAGGTTGTGGCCCAGGGCGTTAAATTCACAGATTTT 3144
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 GTGGAGGAAAGTAGAAATTTACCAAGGTTGTGGCCCAGGGCGTTAAATTCACAGATTTT 307

Qy 3145 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 3204
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308 TTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCAGCACCTTGCT 367

Qy 3205 CTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTTA 3264
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   368 CTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTTA 427

Qy 3265 ATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCTGGAG 3312
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   428 ATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCAGCCTGGAG 475

RESULT 13
US-10-066-543-2949
; Sequence 2949, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2949
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2949

Query Match     9.8%;  Score 333;  DB 15;  Length 472;
Best Local Similarity   99.8%;  Pred. No. 2.2e-159;
Matches  453;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy  2919 ACTTGTCCTAGCGTTTGGAAGAGAGAAAACAGGAACCCACCAATCAACCAA 2978
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ACTTGTCCTAGCGTTTGGAAGAGAGAAAACAGGAACCCACCAATCAACCAA 60

Qy  2979 ACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTGTCATTTTGGTGTATAA 3038
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACAAAGAAAAAATTCCACAATGAAAGAATGTATTTTGTCTTTTGTCATTTTGGTGTATAA 120

Qy  3039 GCCATCAATATTCAGCAAAATGATTCCTTTCTT-AAAAAAAAATGTGGAGGAAAGTA 3097
         ||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   121 GCCATCAATATTCAGCAAAATGATTCCTTTCTTTCTTTAAAAAAAAATGTGGAGGAAAGTA 180

Qy  3098 GAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAA 3157
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAA 240

Qy  3158 ACACAGAAGAAGCTACCTCAGGTGTTTTACCTCAGCACCTTGCTCTTGTTTCCCT 3217
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACACAGAAGAAGCTACCTCAGGTGTTTTACCTCAGCACCTTGCTCTTGTTTCCCT 300

Qy  3218 TAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTT 3277
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTT 360

Qy  3278 GGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACA 3337
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACA 420

Qy  3338 GCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC 3371
         ||||||||||||||||||||||||||||||||
Db   421 GCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACC 454


RESULT 14
US-09-777-564-1124
; Sequence 1124, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSEQ for Window Version 4.0
; SEQ ID NO 1124
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1124

Query Match     7.7%;  Score 264;  DB 9;  Length 264;
Best Local Similarity  100.0%;  Pred. No. 4.9e-124;
Matches  264;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy  3075 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATT 3134
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATT 60

Qy  3135 CACAGATTTTTTAACGAGAGAAAACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCA 3194
         ||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    61 CACAGATTTTTTAACGAGAGAAAACACCAGAAGAAGCTACCTCAGGTGTTTTTACCTCA 120

Qy  3195 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTT 3254
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTT 180

Qy  3255 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 3314
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 240

Qy  3315 GGTGACAGTCCAAGTGTGCAACAG 3338
         ||||||||||||||||||||||||
Db   241 GGTGACAGTCCAAGTGTGCAACAG 264


RESULT 15
US-10-015-219-1124
; Sequence 1124, Application US/10015219
; Publication No. US2003000829A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-015-219-1124

Query Match     7.7%;  Score 264;  DB 15;  Length 264;
Best Local Similarity  100.0%;  Pred. No. 4.9e-124;
Matches  264;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy  3075 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATT 3134
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTGTTGGCCCAGGGCGTTAAATT 60

Qy  3135 CACAGATTTTTTAACGAGAGAAAACACAGAAGAAGCTACCTCAGGTGTTTTTACCTCA 3194
         ||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    61 CACAGATTTTTTAACGAGAGAAAACACCAGAAGAAGCTACCTCAGGTGTTTTTACCTCA 120

Qy  3195 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTT 3254
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCACCTTGCTCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTT 180

Qy  3255 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 3314
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCAGCCTGGAGAA 240

Qy  3315 GGTGACAGTCCAAGTGTGCAACAG 3338
         ||||||||||||||||||||||||
Db   241 GGTGACAGTCCAAGTGTGCAACAG 264


Search completed: July 21, 2004, 23:22:09
Job time : 1438 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2004, 10:30:59 ; Search time 53.5 Seconds
                                          (without alignments)
                                          12269.364 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 6121
Sequence: 1 ggcagcgcggaggacggcgagga.........aaccctgaaaatgttattt 3412

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext   0.5
               Ygapop 10.0 , Ygapext   0.5
               Fgapop   6.0 , Fgapext   7.0
               Delop   6.0 , Delext   7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_16072004_113050_13341/app_query.fasta_1.3591
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437 @CGN_1_1_86 @runat_16072004_113050_13341 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=ICO -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   PIR_78:*
               1:_pir1:*
               2:_pir2:*
               3:_pir3:*
               4:_pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	494	8.1	854	2	T23837	hypothetical prote
2	279.5	4.6	568	2	T49962	hypothetical prote
3	277.5	4.5	398	2	T41600	probable pre-mRNA
4	269	4.4	621	2	D96554	hypothetical prote
5	265.5	4.3	680	2	T25832	hypothetical prote
6	259	4.2	768	2	T27855	hypothetical prote
7	254.5	4.2	589	2	T19216	hypothetical prote
8	254.5	4.2	611	2	T19217	hypothetical prote
9	244	4.0	644	2	A53184	myc far upstream e
10	238.5	3.9	356	2	S58529	alpha-complex prot
11	236.5	3.9	846	2	T04533	hypothetical prote
12	235	3.8	413	2	S46109	hnRNP complex prot
13	226	3.7	479	2	C86275	hypothetical prote
14	226	3.7	510	2	I38489	onconeural ventral

15	225.5	3.7	365	2	S42471	hnRNP protein E2 -
16	225	3.7	362	2	S78515	single-stranded nu
17	223.5	3.6	632	2	T06627	hypothetical prote
18	219.5	3.6	313	2	T48439	probable RNA-bindi
19	217.5	3.6	397	2	T30168	hypothetical prote
20	209.5	3.4	396	2	S41224	hnRNP protein - Af
21	203.5	3.3	1268	2	A44125	high density lipop
22	201.5	3.3	1270	2	S23464	vigilin - chicken
23	198	3.2	1279	2	T41389	rna binding protei
24	194.5	3.2	748	2	T04401	hypothetical prote
25	193.5	3.2	1220	2	T19117	hypothetical prote
26	192.5	3.1	1198	2	B88279	protein C08H9.2 [i
27	186.5	3.0	464	2	S43363	transformation upr
28	186.5	3.0	464	2	A54143	kappa-B motif-bind
29	185.5	3.0	463	2	S41495	dC stretch-binding
30	185.5	3.0	1684	2	JW0057	gravin - human
31	184.5	3.0	629	2	T05425	polyadenylate-bind
32	184.5	3.0	806	2	T13690	hypothetical prote
33	182	2.9	968	2	T00353	hypothetical prote
34	179	2.9	924	2	S27923	gene LF3 protein -
35	179	2.9	1283	2	T49692	related to SCP160
36	179	2.9	4957	2	T03455	ALR protein - huma
37	179	2.9	5262	2	T03454	ALR protein - huma
38	178.5	2.9	473	2	A54691	octamer-binding pr
39	177.5	2.9	471	2	S41768	splicing factor ho
40	176.5	2.9	471	2	G01211	54 kDa protein - h
41	176	2.9	1222	2	S56000	SCP160 protein - y
42	175.5	2.9	1791	2	T02345	hypothetical prote
43	174.5	2.9	817	2	S51342	verprolin - yeast
44	168.5	2.8	2167	2	T30826	nascent polypeptid
45	167	2.7	638	2	S37085	polyadenylate-bind

ALIGNMENTS

RESULT 1
T23837
hypothetical protein M88.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23837
R;Sulston, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19806
A;Accession: T23837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-854 <WIL>
A;Cross-references: EMBL:Z34802; PIDN:CAA84338.1; GSPDB:GN00021; CBSP:M88.5
A;Experimental source: clone M88
C;Genetics:
A;Gene: CESP:M88.5
A;Map position: 1-854
A;Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1

Alignment Scores:
Pred. No.:   6.67e-26   Length: 854
Score: 494.00   Matches: 161
Percent Similarity: 41.86%   Conservative: 96
Best Local Similarity: 26.22%   Mismatches: 182
Query Match: 8.07%   Indels: 176
DB: Gaps: 21

US-09-270-437D-6 (1-3412) x T23837 (1-854)

Qy   422   CAGAAACCGGCCGTTGTCAACGTCACATATGCAACAAGA-------------------- 459
                ||: || ::   :||:
Db   244   GlnHisProGlnMetMetGlnMetGlnHisAlaGlnGlnGlyTyrHisProHisGlnGlnAsn 263

Qy   460   GAAGAAGCAAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTC 519
                ::::: :: :: || ::: |||:||| :: ||||||
Db   264   GlnGlnHisGlnAlaGlyGlnHisGlnSerHisHisGlnSerGlnAsnHisAsn------ 282

RESULT 2
T49962
hypothetical protein F8M21.160 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49962
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T49962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-568 <BEV>
A;Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A;Experimental source: cultivar Columbia; BAC clone F8M21
C;Genetics:
C;Gene: ATSP:F8M21.160
A;Map position: 5
A;Introns: 200/3; 337/3; 544/3

Alignment Scores:
Pred. No.:
Score: 3.52e-11   Length: 568
                                279.50   Matches: 140
Percent Similarity: 40.82%   Conservative: 100
Best Local Similarity: 23.81%   Mismatches: 242
Query Match: 2   Indels: 107
DB: 200/3; 337/3; 544/3   Gaps: 24

US-09-270-437D-6 (1-3412) x T49962 (1-568)

```
Qy   535  CCGATGAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCGTGGGGACCACTCT       594
               |||||      ::             |||||
Db     3  ProAspHisArgMet--------SerProAspHisArgSerHisArgLysArgSer       19

Qy   595  TCCCGGGAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGA---------       639
               :::           |||||  |||:::  |||
Db    20  ArgProGlnSerAspTyrArgAspAsnGlyLysSerLysArgArgTyrArgArgGlyAspAsp   39

Qy   640  --------------CAGATTGATTCCCGCTCGCGGATCCTGGTCCCCACCCAG       678
                           :::      |||||||
Db    40  ArgAspSerLeuValIleAspArgArgAspAspThrValPheArgTyrLeuCysProValLys   59

Qy   679  TTTGTTGGTGCCATCGGAAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGGCAGACC   738
               :::::|||||::||||||:::       |||      :::
Db    60  LysIleLysSerValIleGlyLysArgGlyGlyAlaIleLeuLysGlnLeuArgAsnAspThr  79

Qy   739  CAGTCCCGGGTAGATATCCATAGAAAAGAGAAACTCGAGCTGCAGAGAAGCCTGTCACC    798
               ::::::|||  :::|||  |||      |||||||  :::|||  :::|||
Db    80  ArgSerLysIleArgIle--GlyGluAlaIleProGlyCysAspGlyArgValIleThr     98

Qy   799  ATCCATGCCACCCCCAGAGGGGACT---------TCTGAA                       828
               ::::::::  :::       |||
Db    99  IleTyrSerProSerAspGluThrAsnAlaPheGlyAspGlyGlyLysValLeuSerPro    118

Qy   829  GCATGCGCCATGATCTTGAAATCATGCAGAAAAGAGGCCAGATGAGAACCAGCGGAA     888
               |||      :::      |||:::       |||      :::
Db   119  AlaGlnAsnAspAlaLeuPheArgIleHisAspArgArgValAlaAlaAspAlaArgSerGlu  138

Qy   889  GAGATTCT--------------CTGAAAAATTTGGCACACAATGCCTTGGTT           927
               |||               :::      |||||:::
Db   139  AspSerProGluGlyLysLysGlnValThrAlaLysLeuLeuValProSerAspGlnIle    158

Qy   928  GGAAGACTGATTGGAAAAAGAGGCAGAGAAATTGAAGAAAATGAACATGAAAACAGGGACC   987
               ::::|||:::|||||:::      :::       ||||||
Db   159  GlyCysIleLeuGlyLysArgGlyGlyGlyIleIleValGlnIleLeuArgSerGluThrGlyAla  178

Qy   988  AAGATAACAATCTCATCTTTGCAGGATTTGAGCATA---------TACAACCCGGAAAGAACC  1041
               :::      ::   :::       |||
Db   179  GlnIleIleArgIleValLysValAspAspMetProLeuCysAlaLeuAsnSerAspGluLeu   198

Qy  1042  ATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATTATGAAGAAG        1101
               |||||      :::       ||||||
Db   199  IleGlnIleLeuSerGlyValGluValLeuLysLysSerAlaLeuLeuIleAlaSerArg    218

Qy  1102  CTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAATCTGATCCCA     1161
               |||||      :::::
Db   219  LeuHisGlu--------------AsnProSerArgSerGlnAsnLeu--------        229

Qy  1162  GGGTTGAACCTTCAGGCGCCACTTGGCATCTTTTTCAACAGGACTGTCCGTGCTCTCCACCA   1221
                          ::::     ||||||:::
Db   230  --------LeuSerSerGlyLysTyrProAlaGly---SerLeuMetSerHisAla       245

Qy  1222  GCAGGGCCCCCCGGCAGCTCCCCCGGTGCCCCCTTCACTACCACCTCCGGA          1281
               ||||||||
Db   246  GlyGlyProArgLeuValGlyLeuAlaProLeuMetGlySerTyrArgAspAlaGly      265

Qy  1282  TACTTCTCCAGCCTGTACCCCCGTACCCCAGTTTGGCCCGTTCCCCATCATCACTCCTTAT   1341
               ::::||      :::      |||
Db   266  AspTrpSerArg--------ProLeuTyrGln--------ProProArgAsnAspPro    279

Qy  1342  CCAGAGCCAGGAG---ATTGTGAATCTCTTCATCCCAACCCAGGTGTGGGCGCCATCATC   1398
               ||||||:::       :::      :::
Db   280  ProAlaThrGluPhePheIleArgLeuValSerProValGluAsnIleAlaSerValIle   299

Qy  1399  GGGGAAGAGGGGGCACAGCACATCAAAACAAGCTGCGGAGATTCGGCGGAGCCCTATCAAGATT  1458
               ||||||
Db   300  GlyLysGlyGlyAlaGlnLeuAsnGlnLeuAsnLeuArgGlnGluThrArgAlaThrIleLysVal 319

Qy  1459  -----GCCCCTGCGGAAGGCCAGAC-----------GTCAGCGAAAGGATGGTCATC      1500
                    :::      |||||||
Db   320  AspSerSerArgGlyThrGluArgLysAspCysAsnAspCysSerLeuIleThrIleSerAlaArgGlyValPhe  339

Qy  1501  --------ATCACCGGGCCACCGAAGCCCAGTTCAGGCCCGGATCTTTGGG            1551
```

```
Qy   340  GluAspAlaTyrArgSerProThrIleGluAlaValMetArgLeuGlnProLysCysSerAsp  359
                                    ||||||
Db  1552  AAACTGAAAGAGAAAAACTTCTTTAACCCCAAAGAAGAAGAAGTGAAGCTGAAGGCCATATC    1611

Qy   360  LysValGluArgAspSerGlyLeu--------ValSerPheThrArgLeu           374
               :::                  ||||||
Db  1612  AGAGTGCCCTCTTCCACAGCTGCGGATCCTGGTCAAGGTGGCAAAGGTGAACGAA       1671

Qy   375  LeuValProSerSerArgIleGlyCysIleLeuGlyLysGlyAlaIleIleThrGlu     394
               ::::||||||||||||||
Db  1672  CTGCAGAACTTAACCAGTGCAGAAGTC---ATCGTGCCTCGTGACCAACGCCA--------  1722

Qy   395  MetArgArgMetThrLysLeuAlaAsnIleArgGluLysLysGluLeuProLysLysVal    414
               :::      |||
Db  1723  --------GATGAAAATGAGGAGGAGTGATCGTCAGAATTATCGGGACTCTTCTTGCT       1770

Qy   415  AlaSerAspAspAlaGluMetValGlnValAsnPheMetValLeuLeuLysPheSer      434
               ||||||       :::
Db  1771  AGCCAGACTGCCACAGCGCCAGATCAGGGAAAATTGTACACACAGTGAGCAGCAGGAGGACA   1829

Qy   435  LeuGlnPheLeuLeuSerArgLeuArgTyrPheThrTrpSerValAsnSerAsnSerAsn    454
               :::::::
Db  1830  --GAAATACCCTCAGGGAGTCGCCTCACAGCGCCAGCAAGTGAGGCTCCCACAGGCCACCA    1886

Qy   455  PheAspValGlnIleSerGlyGluLeuAspLysValAlaLysGluAlaLeuIleThr      474
               ::::::
Db  1887  GCAAAACACGGATGAAGTGA                                           1910

Qy   475  SerArgLeuArgAlaAsnValPheAspArgGluGlyGlyAlaValSerArgAlaLeuMetProVal 494
               :::::
Db  1911  CTTCCA---ACACCTGACAGAATGAGACCACAAACGCAGATGGGAGCCAAACCA          1967

Qy   495  LeuProTyrValPro--------ValAlaProAlaGlyAsp-ArgPheAspTyrAspSe   512
               |||:::
Db  1968  AAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGGCCCAGGGGACTCTGCGGAGGCCCTGA   2027

Qy   512  rArgAspSerArgLeuGlu--------ArgGlyAsnPr                          523
                                       |||
Db  2028  GAACCCCAGGGGCGGCCGAGGGGGGCCGGGGAAGGTCAGCCGGTTTGCCAGAACCACCGAGC   2087

Qy   523  oTyrProGlyPheGlyTyrGlySerSerGly--------ValSerAlaGluGlyTyrSe   539
               |||:::                  :::
Db  2088  CCGGCCTCCCGCCCCCCCCAGGGC                                        2109

Qy   539  rProTyrGlyAlaProValGly   546
```

RESULT 3
T41600
probable pre-mRNA splicing factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41600
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z22003
A;Accession: T41600
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-398 <LYN>
A;Cross-references: EMBL:AL031825; PIDN:CAA21234.1; GSPDB:GN00068; SPDB:SPCC757.09c
A;Experimental source: strain 972h-; cosmid c757
C;Genetics:
C;Gene: SPDB:SPCC757.09c
A;Map position: 3

Alignment Scores:

Pred. No.:	4.38e-11	Length:	398
Score:	277.50	Matches:	108
Percent Similarity:	41.38%	Conservative:	60
Best Local Similarity:	26.60%	Mismatches:	157
Query Match:	4.53%	Indels:	81
DB:		Gaps:	15

```
Db   327 IleProAlaAspMetValGlyCysIleIleGlyArgGlyGlySerLysIleSerGluIle 346
Qy  1429 GCCAGATTCGCCGAGCCTCTATCAAGATTGCCCTCGGAAGGCCCAGACGTCAGCGAA 1488
         |||   ||||:||: ||| :::||| ||| ||||  |||
Db   347 ArgArgThrSerGlySerLysIleSerIleAlaLysGluProHisAspGluThrGlyGlu 366
Qy  1489 AGGATGGTCATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTT 1548
         ||||||  |||::  :||| |||| || :: |||||||||| |||:
Db   367 ArgMetPheThrIleThrGlyThrHisGlyGluGluAsnGluLysAlaLeuPheLeuLeuTyr 386
Qy  1549 GGGAAACTGAAAGAGGAA 1566
         :::||||:||
Db   387 GlnGlnLeuGluMetGlu 392
```

RESULT 4
D96554
hypothetical protein F19C24.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96554
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, D.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A8614.; MUID:21016719; PMID:11130712
A;Accession: D96554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-621 <STO>
A;Cross-references: GB:AE005173; NID:g11094762; PIDN:AAG29695.1; GSPDB:GN00141
C;Genetics:
A;Gene: F19C24.19
A;Map position: 1

```
Alignment Scores:
Pred. No.:
Score:                    1.9e-10        Length:               621
                          269.00         Matches:              134
Percent Similarity:       37.37%         Conservative:          88
Best Local Similarity:    22.56%         Mismatches:           206
Query Match:               4.39%         Indels:               166
DB:                       2              Gaps:                  26
                                         (1-621)
```

US-09-270-437D-6 (1-3412) x D96554 (1-621)

```
Qy   616 CCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTGCGGATCCGGTCCCCACC 675
         |||||||||:||:||    :::|||
Db     9 ProAlaThrThrAlaThrAlaAlaGlu--SerValHisPheArgLeuLeuCysProAla 27
Qy   676 CAGTTTGTTGGTGCCATCATCGGAAAGAGAGGGCTTGACCATAAAGAACATCACTAAGGACAG 735
         |||||:||::|  :||| ||||||:::|||
Db    28 ThrArgThrGlyAlaAlaIleIleGluIleGlyLysLysGlySerValIleGlnSerVal 47
Qy   736 ACCCAGTCCCGGGTA 750
         ||||:||:::|||
Db    48 ThrGlySerLysIleArgValIleAspAspIleProValProGluGluArgValVal 67
Qy   751 ---------------GATATCATAGAAAAGAGAAC 771
                        :::||||:||
Db    68 LeuIleIleAlaProSerGlyLysLysLysAspGluSerAsnValCysAspSerGluAsn 87
Qy   772 TCTTGGAGCTGCAGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGGGACT------ 822
         |||:||||||:::|||
Db    88 ProGlySerGluGluProLysGluProLysGlyLysGlySerGluCysAlaGlyThrSerGlyGly 107
Qy   823 -----------TCTGAAGCATGCCGCATGAGATTCTTGAAATCATGCAGAAA--- 861
                    :::||||||:::|||
Db   108 AspAspGluGluAlaProSerSerAlaGlnMetAlaLeuLeuArgValPheArgIle 127
```

US-09-270-437D-6 (1-3412) x T41600 (1-398)

```
Qy   505 GAGAACTACTCCTTCAAGATTCCTAC------ATCCCGATAAGAGAGTGAGCTCC--- 555
         ||||||  :::||| :::||:::
Db    16 GluAsnSerPhePheAspValThrPheGlnAspGluProAspGluThrhrSerThrAla 35
Qy   556 ------------CCTTCGCCCCCT---CAGCGAGCC 576
                     ||||||
Db    36 ThrGlyIleAlaLysValSerIleProThrProLysProLeuSerThrLeu 55
Qy   577 CAGCTGGGGACCACTCTTCCCGGAGACAAGGCCACGCCCTGGGGGCACTTCTCAGGCC 636
         :::     |||  :::|||||
Db    56 ThrAsnGlySerThrIleGlnSerMetThrAsnGlnProGluProThrSerGlnVal 75
Qy   637 AGACAGATT-------------GATTTCCCG 654
         ||||||||
Db    76 ProProIleSerAlaLysProMetAspAspAlaThrTyrAlaThrGlnGlnLeuThr 95
Qy   655 CTGCCGATCCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGAAAGAGAGGGCTTGACC 714
         |||||  |||::  |||||||:||
Db    96 LeuArgAlaLeuLeuSerThrArgGluAlaGlyIleIleIleGluIleGlyLysLysAsn 115
Qy   715 ATAAAGAACATCACTAAGGACAGACCCAGTCCGGGTAGATATCATAGAAAAGAGAACTCT 774
         |||   :::||| :::
Db   116 ValAlaGluLeuArgSerThrArgAsnValLysAlaGlyVal--ThrLysAlaValPro 134
Qy   775 GGAGCTGCAGAGAAGCCTGTCACCATCATGCCGCACCCCAGGAGGGACTTCTGAAGCATGC 834
         |||||:::|||
Db   135 AsnValHisAspArgValLeuThrIleSerGlyProLeuGluAsnValValArgAlaTyr 154
Qy   835 CGCATGAGATTCTTGAAATCATGCAGAAAGAGGCA---------GATGAGACCAAACTAGCC 885
         |||:::|||
Db   155 ArgPheIleIleAspIlePheAlaLysAlaAsnSerThrAsnProAspGlyThrProSerAsp 174
Qy   886 GAAGAGAATTCCT-----CTGAAAATCTTGGCACCACAATGGCTTGGTTGGAAGACTGATT 939
         :::||||:::|||  :::||: :::|||
Db   175 AlaAlaAsnThrProArgGlyLeuArgLeuLeuIleAlaHisSerLeuMetGlySerIleIle 194
Qy   940 GGAAAAGAAGGCAGAGAAATTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATC 999
         :::|||  ::: ::
Db   195 GlyArgAsnGlyLeuArgIleLysLysLeuIleGlnIleAspLysCysSerCysArgMetIleAla 214
Qy  1000 TCATCTTTGCAGGATTGAGCATATACAACACCCGGAAAGAACCATCACTGTGAAGGGCACA 1059
         |:|||  :::|||
Db   215 Ser------LysAspMetLeuProGlnSerThrGlnArgThrValGluIleHisGlyThr 232
Qy  1060 GTTGAGGGCCCTGTCCCAGTGCTGAGATAGAGAATTATGAAGAAGCGCGTGGAGGCCTTTGAA 1119
         |||:::||||:||  :::|||
Db   233 ValAspAsnLeuHisIleAlaAlaIleAlaIleTrpGluIleGlyIleLysCysLeuIleLeuAspAspTrpGlu 252
Qy  1120 AATGATAATGCTGGCTGTTAACCAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCCCA 1179
         :::|||
Db   253 Arg--------------------GlyAla 255
Qy  1180 CTTGGCATCTTTCAACACGTCCGTCTATCTCCACCA------ 1221
         :::||| :::|||
Db   256 GlyThrValPheTyrAsnProValSerArgLeuThrGlnLeuProLeuProSerLeuAlaSer 275
Qy  1222 --GCAGGGCCCCGCGGA--GCTCCCCCGCTGCCCCCTTCACTACCCCAC 1275
           |||:::|| :::|||
Db   276 ThrAlaSerProGlnGlnValSerProProAlaAlaPro--------SerThrThr 291
Qy  1276 TCCGGA----------TACTTCTCCAGCCTGTACCCCCATCAC 1308
         |||||
Db   292 SerGlyGluAlaIleProGluAsnPheValSerTyrGlyAlaGlnValPheProAlaThr 311
Qy  1309 CAGTTTGGCTGCCAGAGAAGCCTGTCACCATCATGCCACCAGCAGGATTGTGAATCTTTC 1368
         |||||:::|||
Db   312 GlnMet---ProPheLeuGlnGlnProLysValThrGln------------AsnIleSer 326
Qy  1369 ATCCCAACCCAGGCTGTGGCGGCCATCATCGGAAGAGAGGGGCCACACATCAAACAGCTG 1428
         |||||  :::|||
Db   108 AspAspGluGluAlaProSerSerAlaGlnMetAlaLeuLeuArgValPheArgIle 127
```

RESULT 5

T25832

hypothetical protein M01A10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R;Scheet, P.
C;Accession: T25832
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid M01A10.
A;Reference number: Z20094
A;Accession: T25832
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;residues: 1-680 <SCH>
A;Cross-references: EMBL:U88174; PIDN:AAB42272.1; GSPDB:GN00019; CESP:M01A10.1
A;Experimental source: strain Bristol N2; clone M01A10
C;Genetics:
A;Gene: CESP:M01A10.1
A;Map position: 1
A;Introns: 63/2; 97/3; 121/1; 160/1; 269/3; 411/3; 512/1; 649/2

Alignment Scores:

Pred. No.:

Score:	265.50
Percent Similarity:	38.30%
Best Local Similarity:	20.45%
Query Match:	4.34%
DB:	

Length:	680
Matches:	118
Conservative:	103
Mismatches:	213
Indels:	143
Gaps:	17

Score: 3.4e-10

US-09-270-437D-6 (1-3412) x T25832 (1-680)

```
Db   138  ArgValAspThrSerThrValIleLysAlaSerValSerIleProGluGluSerVal  157
Qy   685  GGTGCCATCTCGGAAAAGGAGGGGTTGACCATAAAGAACATCACTAAGGCAGACCCAGTCC  744
          ::::||||||||  ||||::: |||:: |||||||||| :::::::
Db   158  GlyLeuValIleGlyArgAsnGlyValGluIleLysSerGlnLysSerGlyLysSerGlyCys  177

Qy   745  CGGGTAGATATCCATAGAAAAACAGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCCAT  804
          ||||||                                ||| ::: ||:::||||||||
Db   178  ArgValGlnIle---ValAlaGluProSerThrThrGlyTyrArgSerValAspIleTyr  196

Qy   805  GCCACCCCAGAGGGACTTCTGAAGCATGCGCATGATCTTGAAATCATGCAGAAAGAG  864
          |||                   :::::||||||| ||:::::
Db   197  GlyIleSerGluAsnIleGluValAlaLysLysLeuIleAsnGluValVal------  213

Qy   865  GCAGATGAGACCAAACTAGCCGAGAG----------------ATTCCT  897
          |||                               ||||||
Db   214  AlaArgGlyArgLysLysSerGlnLeuProCysSerValProGlnPheGlnPro  233

Qy   898  CTGAAAATCTTGGCACACAATGGCTGGTT--  927
          ::: |||||||||::::::
Db   234  IleProAlaValSerAsnSerSerLysValThrIleIleIleIleProIleProAlaAsnLys  253

Qy   928  --GGAAGACTGATTGGAAAAGAAGGCAGAAATTGAAGAAATTGAACATGAAACAGGG  984
          ||| :::::||:::|||
Db   254  CysGlyLysAlaIleLeuGlyLysLysGlyGluGlnLeuMetArgLysLeuArgSerTrpThrAsn  273

Qy   985  ACCAAGATAACAATCTCATCTTTGCAGATTGAGCATTACAACCCGGAAAGAACCATC  1044
          ::|: ||
Db   274  CysAspPheIleLeu------IleGluGluAsnAsnIleAlaAspSerValLysProLeu  291

Qy   1045 ACTGTGAAGGGCACAGTTGAGCGCTGTGCAGTTGCTGAGATAGAGATTATGAGAAGCTG  1104
          ||| ::::||| |||
Db   292  GlnIleThrGlyGlnProLys-----------GluValGluHisAlaLysAlaLysLeu  306

Qy   1105 CGTGAGGGCCTTTGAAAAATGATATGCTGGCTGTTAACCAACAGGCCAATCTGATCCCAGGG  1164
          |||                        ValAlaAspIleLeuAspGly
Db   307  ------------------------------------ValAlaAspIleLeuAspGly  313

Qy   1165 TTGAACCTGCAGCCGCACTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCA  1224
          |||||||||                     ||||||||:::|| ||:::
Db   314  PheAsp----------------------------GluCysProProAla  320

Qy   1225 GGGCCCCGGAGCTCCCCGCTGCCCCCTACCACCCCTTCACTACCTCCGGATAC------  1284
          ||| |||
Db   321  GlyMetAlaGlyAsnSerProValAlaAlaMetSer---  332

Qy   1285 TTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTTCCCGCATCACTCTTATCCA  1344
          |||                                          |||
Db   332  -------------------------------------------------  332

Qy   1345 GAGCAGGAGATTGTGAATCTTCTCATCCCAACCCCCAGGCCCCATCATCGGGAAG  1404
          ||||||||||::::||||
Db   333  ------------------LeuGlnValLysValProArgSerThrValGlyAlaIleMetGlyMetLeu  348

Qy   1405 AAGGGGGACACATCAAACAGCTGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCCCCT  1464
          ::::||||||:::|||||||
Db   349  GlnGlySerAsnIleLysSerLeuSerAsnGluThrGlyThrLysIleGlnPheMetPro  368

Qy   1465 GCGGAAGGCCCAGACGTCAGCGAAAAGGATGGTCATCATCACCGGGCCACCGGAAGCCCAG  1524
          ::: ||||||:::|
Db   369  AspAspAspProLysLeuMetGluArgThrLeuValValIleGlyAsnLysAsnLysSerVal  388

Qy   1525 TTCAAGGCCCAGGGACGGGATCTTTGGGAAAACTGAAAGAGGAAAAACTTCTTTAACCCCAAA  1584
          ::|                    ||||| ||:::
Db   389  TyrValCys--AlaArgLeuLeuGlnLysIleValGluAlaAsnSerGluAsnAlaAsn  407

Qy   1585 GAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCGGGTGATT  1644
          :::|||||||||::::||:::
Db   408  ThrProIleSerLeu--PheTyrMetLeuIleProAlaSerLysSerCysGlyLeuValIle  426

Qy   1645 GGCAAAGGTGGCAAGACGCTGACGAAGAACGTCAGAACTTAACCAGTGACGAAGTCATCGTG  1704
          ||||::::||||||||||:::|||
Db   427  GlyArgGlyGlyGluThrIleArgGlnIleAsnLysAsnLysSerGlySerGluSerGlyAlaTyrCysGluMet  446
```

```
Qy   1705 CCTCGTGAC----  ----CAAACG  1719
          ||||
Db   447  SerArgAspProSerIleSerGluAlaIleAlaIleGluLysGlnPheValIleArgGlySerGluThr  466

Qy   1720 CCAGATGAAATGAGGAAGTGATCGTCAGATTATCGGGACGATTATCCTTTGCTAGCCAGACT  1779
          ||||||||||| :::  ::::::|||
Db   467  GlnIleValGluHisAlaLysHisLeuIleArgVal-----------------  477

Qy   1780 GCACAGGCGCAAGATCAGGGAAATT--  ----GTACAA  1809
          |||:: |||:::
Db   478  ----------LysValGlyAspIleProProAsnThrProTyrIleAsnThrArgAlaAla  494

Qy   1810 CAGGTGAAGCAGCAGGAGCAGAAAATACCCTCAGGGAGTCGCCTCAGCCGCAGCCAAGTGA  1869
          :::  ::: |||:::
Db   495  GlnProLeuGlnPheSerHisGlnAsnProThrAlaIleAspSerTrpArgAlaGlnPro  514

Qy   1870 GGCTCCCACAGGCACCAGCAAAACAACGGATGAAATGTAGCCCTTCCAACACCTGACAGAA  1929
          :::  |||   |||||||
Db   515  PheThrThrGlnHisGlnAsnSerLeuSerProGlnProGlnAlaHisGlnPhePro  534

Qy   1930 TGAGAACCAAACGCAGCCAGCCAGATCGGGACGAACAAACCAAAGACCAATCATCTGAG  1980
          |||||||                  ::: |||
Db   535  AsnLeuMetAlaTyrSerAlaArgLeuGlyTyrGlnSerHisProProGln  551
```

RESULT 6

T27855
hypothetical protein ZK418.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27855
R;Fulton, L.
submitted to the EMBL Data Library, April 1994
A;Description: The sequence of C. elegans cosmid ZK418.
A;Reference number: Z20430
A;Accession: T27855
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-768 <FUL>
A;Cross-references: EMBL:U00047; PIDN:AAA50693.1; CESP:ZK418.9
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:ZK418.9
A;Introns: 166/1; 221/3; 243/1; 279/1; 387/3; 490/1; 567/3; 727/2

Alignment Scores:
Pred. No.:   Length: 768
Score: 9.86e-10   Matches: 134
Percent Similarity: 259.00   Conservative: 86
Best Local Similarity: 38.33%   Mismatches: 218
Query Match: 23.34%   Indels: 136
DB: 4.23%   Gaps: 23
                           2

US-09-270-437D-6 (1-3412) x T27855 (1-768)

```
Qy   560  CGCCCCTC---AGCGAGCCCAGGCGTGGGGACCCACTCTTCCGGGAGCAGGCCACGCCC  616
          ::|||                  |||:: |||
Db   234  ArgProLeuAspSerGluIleLeuLeuAspGlyAspProThrLysLysSerGlu  253

Qy   617  CTGGGGGCACTT---CTCAGGCGCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCA  673
          ::||| |||                  |||
Db   254  ValGlyAspLeuAsnMet-GlyAspSerAspLysIleThrAspIleTyrProG-ValProG  273

Qy   674  CCCAGTTGTTGTTGGTGCCATCATCGGAAAAGAGGGGCTTGACCATAAAGAACATCACTAAGGC  733
          :::||||                   ||||:::
Db   273  luLysValValGlyValLeuValIleGlyLysValGlyLysValSerGluValIleLeuGlnIleT  293

Qy   734  AGAGCCCAGTCCCGGGTAGATATCCATAGAAAAAGAGAAGAGCTG  793
          :::                        |||:::
Db   293  hrSerGlyCysArgValGlnMetAspProAspHisGlnSerValAsnGlyPheArgAsnC  313

Qy   794  TCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATCA  853
          |||||                      :::|||
```

```
Db   313 ysThrIleGluGlyProProAspGlnValAlaValAlaAlaArgGlnMetIleThrGlnValI 333

Qy   854 TG---------CAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAA 904
         ::   ||||| |||
Db   333 leAsnArgAsnGlnThrGlyAlaGlnProGlyAlaAlaProGlyValThrGluGluM 353

Qy   905 TCTTGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTGAAGA 964
         :::||| |||||||| ::::|||
Db   353 etLeuIleProAlaAspLysIleGlyLeuValIleGlyLysLysGlyGluThrIleArgI 373

Qy   965 AAATTGAACAATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCATAT 1024
         ||| ||||    :::::||
Db   373 leValGlnGluGlnSerGly---------------LeuArgAsnCysAsnValV 386

Qy  1025 ACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGT---------G 1072
         :::  |||   |||
Db   386 al------GlnGluThrThrAlaThrGlyAlaThrGlyGlnProLysProLeuArgMetIleGlyLys 404

Qy  1073 CCAGTGCTGAGATAGAGATTATGAAGAGCTGCGTGAGCGCTTTGAAAATGATATGCTGG 1132
         ||||||::: |||
Db   404 erProAlaAlaIleGluThrAlaLysAlaLeuValHisAsnIleMetAsn---------- 420

Qy  1133 CTGTTAACCAACAAGCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTT 1192
         |||                   ||||| :::::|||
Db   421 -----AsnThrGlnGlyAsn- 425

Qy  1193 CAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCC------CCCG 1246
         ||||||
Db   426 ----------------------------------AlaProLeuLeuGlnA 431

Qy  1247 CTGCCCCTACCACCCCTTCACTACCCACTCCGATACTTCTCCAGCTCTGTACCCCCATC 1306
         ||||||:::
Db   431 rgAlaProHisGlnPro----------SerGlyGlnPheGly-- 441

Qy  1307 ACCAGTTTGGCCCGTTCCCGCATCATCACTCTTATCCAGCAGGAGATTGTGAATCTCT 1366
         |||                  |||:::||||
Db   442 ---------GlyGlyTyrGlyAlaGlyInGluAlaGlnLysAlaLysGlyGluValIle-- 456

Qy  1367 TCATCCCAACCCAGCGTGTGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAAACAGC 1426
         :::|| :::  |||:::   ||||||
Db   457 --ValProArgLeuSerAlaGlyMetIleIleIleGlyLysLysGlyGluMetIleLysLysArgL 476

Qy  1427 TGGCCAGATTCGCCGGAGCTCTATCAAGATTGCCCCTCGCGAAGGCCGACGTCAGCG 1486
         |||                      :::|| :::
Db   476 euAlaAlaGluThrGlyAlaThrIleArgAsnProAspThrAsnProAsnSerGluA 496

Qy  1487 AAAGGATGGTCATCATCACCGGGCCACCCGGAAGCCCAGTTCAAGGACGGAGGACGATCT 1546
         :::||::: :::                        |||
Db   496 spArgIleAlaValIleMetGlyLysLysAspThrArgAspArgAlaThrArgGlyIleT 516

Qy  1547 TTGGGAAACTGAAAGAG---------GAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGG 1600
         :::                        :::::|||
Db   516 hrGlyIleValAsnArgAlaIleLeuLysAsnAsnGlyAlaProGlnAspAspArgGlySerAlaG 536

Qy  1601 AAGCG------------------CATATCAGAGTGCCCTCTTCCACAGCTGGCC 1636
         :::                     ::::|||
Db   536 lyThrValLeuProGlyGlnSerIlePheTyrMetHisValProAlaGlyLysLysCysGlyL 556

Qy  1637 GGGTGATTGGCAAAGTGGCAAGACCGTGCAGACGACTGAACCAGTGCAGAAG 1696
         ||||||:::
Db   556 euValIleLysGlyGluGlyAlaSerIleLysGlnIleLysGluThrGlyAlaThrC 576

Qy  1697 TC---ATCGTGCCTCGTGACCAAACGCCAGATGAGAAATAGGAGAAGTGATCGTGCAGAATTA 1753
         :::                                   ||||||
Db   576 ysGlyIleuLeuAlaProAlaAlaGluGluGlnLysAsnGlnLysGluLysProValPhe---GluIleL 595

Qy  1754 TCGGGCAC-----------TTCTTTGCTAGCCAAGTGGACTGCACAGCCGCCAGTGACTCAGGAA 1801
         :::||||||
Db   595 ysGlySerGlnLeuGlnIleHisHisAlaSerHisLeuValArgIleLysValGlyGluI 615

Qy  1802 TTGTACAACAGGTG----------- 1825
         ||
Db   615 leSerProAsnThrProValProProLeuGlnGlyLeuGlyGlyTyrGlyGlyGlnGlyGlnGlnG 635
```

```
Qy  1826 AGCAGAAATACCCTCAGGGAGTC------GCCTCAGCGCGAGCGAGCAAGTGAGGCTCCCACA 1879
         :::||||| |||
Db   635 lnAlaMetPheSerAlaGlyInThrGlnAlaGlyAlaGlyValTyrGlyValGlySerThrGlyPheIleG 655

Qy  1880 GGCACCAGCAAAACAACGATGAATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAA 1939
         :::||||||           GlnPheGlnHis |||
Db   655 lnGlnGlnGlnInPro-----------GlnGlnG 667

Qy  1940 CGGCAGCCAGCCAGATGCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAGTCTGCGGAG 1999
         ||| |||
Db   667 lnGlnInTrpAla--- 670

Qy  2000 GCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCAGGGGCCGAGGGGGGGCGAGGGGGAAGG 2059
         :::|||::::                   GlyGlyGlyValValGlnG 679
Db   671 ProGlnAsn-----------------------
```

RESULT 7
T19216
hypothetical protein C12D8.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19216
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: 219092
A;Accession: T19216
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-589 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98232.1; GSPDB:GN00023; CESP:C12D8.1a
A;Experimental source: clone C12D8
C;Genetics:
A;Gene: CESP:C12D8.1a
A;Map position: 5
A;Introns: 7/3; 25/1; 159/3; 318/3; 513/2

Alignment Scores:

Pred. No.:		Length:	589
Score:	254.50	Matches:	118
Percent Similarity:	38.31%	Conservative:	82
Best Local Similarity:	22.61%	Mismatches:	180
Query Match:	4.16%	Indels:	142
DB:	2	Gaps:	19

US-09-270-437D-6 (1-3412) x T19216 (1-589)

```
Qy   667 GTCCCACCCAGTTTGTTGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAACATC 726
         :::||| |||||||:::   ||||||||||||||| :::|||||
Db    54 IleProGluSerAlaValAlaValIleLylevalIleGlyLysLysGlyGluGlnIleIle 73

Qy   727 ACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAGAAAGAACTCTGGAGCTGCAGAG 786
         :::                    ::: |||
Db    74 GlnAlaLysLysAlaGlyLysCysArgValGlnMetSerProAlaAspProSerGlyVal 93

Qy   787 AAGGCTGTCGACCATCCATGCCACCCCAGGAGGGGACTTCTGAAGCATGCCCATGACGAATTCTT 846
         ||| :::                  |||||||||
Db    94 ArgMetValThrLeuGluGlulyGluGlnTyrValGluArgSerAsnValGluThrAlaLysHisIleLeuAsn 113

Qy   847 GAAATCATGCAGAAAGAGGCAGAT---------GAGACC 876
         :::::|| :::  |||:::
Db   114 GluValValAlaArgSerGlnAsnProArgProInTyrGlyPheProArgAlaGlnThr 133

Qy   877 AAACTAGCCGAAGAGATTCCTGAAAATCTTGCACAACATCTTGGTTGGAAGACTG 936
         |||
```

```
Qy  1954  TCG------GGAGCAAACCAAAGACCATCTGAGGAATGAGAGTCTGCGGAGGAGGCGGCCAG  2007
                     ::: :::      |||      |||:::
Db   413  AsnAsnPheGlyGlnGlnGlnTrpAsnProValAlaGlnIleProAlaAlaAlaGln  432

Qy  2008  GGACTCTGCCGAGGCCCTGAGACCCCAGGGCGCCGAGGGGCGGGGAAGTCAGCAG  2067
                                          |||||
Db   433  -----------------------AsnProTyrGlnValGlyGlyTrpGlnGlnAsnSerVal  445

Qy  2068  GTTTGCAGAACCACCAGCCCGCCTCCCCGCCCCCCAGGGCCTTCTGCAGGCTTCAGCCA  2127
                                  |||:::   ||:::|||
Db   446  TyrAlaGlnGlnThrAlaAlaProAlaAlaAlaAlaProTyr-AlaAlaAlaGlyIleValGl  465

Qy  2128  TCCA  2131
          |||
Db   465  nPro  466
```

RESULT 8
T19217

hypothetical protein C12D8.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19217
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19092
A;Accession: T19217
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-611 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98233.1; GSPDB:GN00023; CESP:C12D8.1b
A;Experimental source: clone C12D8
C;Genetics:
C;Gene: CESP:C12D8.1b
A;Map position: 5
A;Introns: 181/3; 340/3; 535/2

Alignment Scores:
Pred. No.:   1.89e-09   Length: 611
Score: 254.50   Matches: 118
Percent Similarity: 38.31%   Conservative: 82
Best Local Similarity: 22.61%   Mismatches: 180
Query Match: 4.16%   Indels: 142
DB: 2   Gaps: 19

US-09-270-437D-6 (1-3412) x T19217 (1-611)

Qy   667 GTCCCCACCCAGTTGTTGGTGCCATCATCGGAAGAGGAGGCTTGACCATAAAGAACATC 726
          :::|||| |||||| ||||:::::::
Db   76 IleProGluSerAlaValGlyIleValIleGlyIleValGlySerGluIleGlnGlyIle 95

Qy   727 ACTAAGCAGACCAGTTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTCCAGAG 786
          :::||||| |||| :::
Db   96 GlnAlaLysAlaLysGlyCysArgValGlnMetSerProAlaAspProSerGlyValVal 115

Qy   787 AAGCCTGTCACCATCCATGCCACCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTT 846
                                             |||||| :::
Db   116 ArgMetValThrLeuLysIleSerArgSerArgSerAsnValGluThrAlaLysHisLeuAsn 135

Qy   847 GAAAATCATCGCAGAAAGAGGAGGCAGAT-------------------GAGACC 876
          :::||| :::
Db   136 GluValValAlaArgSerGlnAsnProArgProGlnTyrGlyPheProArgAlaGlnThr 155

Qy   877 AAACTAGCCGAAGAGAGATTCCTCTGAAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTG 936
          :::   :::
Db   156 ThrIleAspIleAlaIlePro-----------------------ProAsnArgCysGlyLeuIle 169

Qy   937 ATTGGAAAAGAAGGCAGAAATTGAAGAAAAATTGAACATGAAAACAGGACCAAGATAACA 996
          |||||| ::::::::::: :::::|| |||::
Db   170 IleGlyLysSerGlyAspThrIleArgGlnLeuGlnLeuLysSerGlyLysCysLysMetIle 189

Qy   997 ATCTCATCTTTGCAGGAT---TTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAG 1053
          |||||| :::|||
Db   190 GlnThrGlnPheSerAlaProAlaGln 412
```

```
Qy  2068  GTTTGCCAGAACACCGAGCCCGCCTCCGCCCCCAGGGGTTCTGCGAGGTTCAGCCA  2127
                |||
Db   468  TyrAlaGlnThrAlaAlaAlaProAlaAlaAlaProTyr-AlaAlaAlaGlyIleValGl  487

Qy  2128  TCCA  2131
                |||
Db   487  nPro  488
```

RESULT 9
A53184
myc far upstream element-binding protein - human
N;Alternate names: FUSE-binding protein
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: A53184
R;Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Krutzsch, H.; Avigan, M.; Levens
Genes Dev. 8, 465-480, 1994
A;Title: A sequence-specific, single-strand binding protein activates the far upstream
A;Reference number: A53184; MUID:94170991; PMID:8125259
A;Accession: A53184
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-644 <DUN>
A;Cross-references: GB:U05040; NID:g460151; PIDN:AAA17976.1; PID:g460152
C;Keywords: DNA binding

Alignment Scores:
Pred. No.:   1.01e-08
Score: 244.00
Percent Similarity: 35.64%
Best Local Similarity: 22.76%
Query Match: 3.99%
DB: 2

Length: 644
Matches: 145
Conservative: 82
Mismatches: 241
Indels: 169
Gaps: 27

US-09-270-437D-6 (1-3412) x A53184 (1-644)

```
Qy   336  CCCTCCTCCACCTGCGAGTGGGAGGTGTTGATGGACTTTTGGCTCAATATGGACAGTGGA  395
                |||||||||| ::::                              ||||||
Db    10  ProSerSerGlySerAlaGlyGlyGlyGly-------------------GlyGlyGly  22

Qy   396  GAATGTGAACAAGTCAACACAGACAGAAACGCCGTTGTCAACGTCACATATGCAAC  455
                ::: ||||||| |||:::                         -----AlaAr
Db    23  GlyGlyGly-GlyValAsnAspAlaPheLysAspAlaLeuGlnArg----------AlaAr  39

Qy   456  AAGGAGAAGCAAAAATAGCCATGGAGAAGCTAAGCGGCATCAGTTTGAGAACTACTC  515
                :::                 ||||||    :::
Db    39  gGlnIleAlaAlaLysIleGlyGlyValGlyThrSerLeuAsnSerAsnAspTyrgl  59

Qy   516  CTTCAAGATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGC  575
                ||||                      |||:::
Db    59  yTyrGlyGlyGlnLysValProLeuGluAspGlyAspGlnProAspAlaLysLysValAl  79

Qy   576  CCAGCGTGGGAGACCACTCTTCCCGGAGCAAGGCCACCGCCCCTGGGAGCCACTCTTCAGGC  635
                ||::                                  ::
Db    79  aProGlnAsnAspSerPheGlyThrGlnLeu------ProProMetHisGlnGlnGlnGlnSe  97

Qy   636  CAGACAGATTGATTTCCCGTGCGGATCCTGGCTCCCACCCAGTTTGTTGGTGCCCATCAT  695
                :||| ||||                       ||||||
Db    97  rArgSerVal--MetThrGluTyrLysValProAspGlyMetValGlyPheIleIl  116

Qy   696  CGGAAAGGACGGCTTGACCATAAAGACACATCACTAAGCAGACCCAGTCCCGGGTAGATAT  755
                |||||::                        :::::
Db   116  eGlyArgGlyGlyGluGlnIleSerArgIleGlnIleSerGlyCysLysIleGlnII  136

Qy   756  CCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCATGCCACCCCAGA  815
                ||| |||||                          ::
Db   136  e---AlaProAspSerGlyLeuProGluArgSerCysMetLeuThrGlyThrProGl  155

Qy   816  GGGGACTTCTGCAAGCATGCCGAGAGGGGCGGGGAAGGTCGCCAG  861
                ||||||                         :::|||
Db   155  uSerValGlnSerAlaLysArgLeuLeuAspGlnIleValGlyGlyArgProAlaPr  175
```

```
Db   190  Leu------ValGlnAspAsnGlnSerValSerAspGlnSerLysProLeuArgIleThr  207
Qy  1054  GGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGC  1113
                ||| :::        :::||||||||::   ||||||:::
Db   208  GlyAspProGln-------LysIleGluLeuAlaLysGlnLeu---------  219
Qy  1114  TTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTC  1173
                                                   :::      ||||||
Db   220  ---------------ValAlaGluIleLeuAsnSer  226
Qy  1174  AGCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGC  1233
                ||||||                   ::: ||||||
Db   227  GlyGlyAspGlyAsnGlyGlySerGlyLeuGlnMet---------  238
Qy  1234  GGAGCTCCCCCGCTGCCCCTACCACCCCTTCACTACCCACTCCGGATACTTCTCCAGC  1293
                -------                                          :::
Db   238  ---------------  238
Qy  1294  CTGTACCCCCATCACCAGTTGGCCCGTTCCGGATCATCACTCTTATCCAGAGCAGGAG  1353
                ||||||                       :::
Db   239  --------HisHisAlaGlyGly--------GlyGlyValAlaSerAlaArgGlyGlu  252
Qy  1354  ATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGCGCCCATCATCGGGAGAAGGGGGGCA  1413
                :::||| |||||        |||||||||:::|||||||||:::
Db   253  ValVal--------ValProArgSerValGlyIleIleIleGlyLysGlyGlnGlyAsp  269
Qy  1414  CACATCAAACAGCTGCGAGATTCGCGGAGCCCTCTATCAAGATTGCCCCTGCCGAAGGC  1473
                ||||||:::||||||                  |||    :::
Db   270  ThrIleLysArgLeuAlaMetGluThrGlyThrLysIleGlnPheLysGlnProAspAspAsp  289
Qy  1474  CCAGACGTCAGCGAAAGGATGGTCATCATCACCCGGGCCACCGGAGCCCAGTTCAAGGCC  1533
                ||||||:::|||||                 ::: ||||||
Db   290  ProSerThrProGluArgCysArgValAileIleMetGlyThrArgAspGlnLysValArgAla  309
Qy  1534  CAGGGACGGACGATCTTTGGGAAACTGAAAGAGGGAAAACTTCTTTAAC----------  1578
                ||||||:::                       |||||
Db   310  ThrGluArgIleThrGluThrLysLysSerThrMetGlnIleGlnLeuGlyGlyGlyAsn  329
Qy  1579  ---CCCAAAGAAGAGTGAAGCTGGAAGCG--------CATATCAGAGTGCCCTCTTCC  1626
                :::||||||                   |||||| ::::::
Db   330  ValAlaLeuGlyAlaMetValSerAsnGluAlaSerThrPheTyrMetSerValProAlaAla  349
Qy  1627  ACAGTTGGCCGGTGATTGGCAAAGTGGCAAGACCGTGAAGACGAACTTAACC  1686
                |||    ||||||::: ||| ::: |||
Db   350  LysCysGlyLeuValIleGluGlyLysGlyGlyGluThrIleLysLysIleGlnAsnSerGlnSer  369
Qy  1687  AGTGCAGAAGTCATCGTCCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTC  1746
                :::                 |||::||||||:::|||
Db   370  GlyAlaHisCysGluLeuSerArgAspProThrGlyAsnAlaAspGluLysValPheVal  389
Qy  1747  -----------------AGAATTATCGGGCACTTCTTTGCTGCAGCCCAGACTGCACAGCCCAG  1791
                                      |||  ::: ||||||
Db   390  IleLysGlyLysLysArgAlaIleGluHis-------AlaLysHisLeuIleLeuArgIleLys  407
Qy  1792  ATCAGGGAAATTGTACAACAGTGAGCAGCAGGAGAAATACCCTCAGGGAGTCGCC  1851
                ::: :::|||                          ||||||
Db   408  ValGlyAspIleAlaProAsnThr---------ProProMetHisGlnGlnGlnSe  415
Qy  1852  TCACAGCGCCAGCAAGTGAGGGTCCCACAGGACCACAGCAAAAACAACGGATGAATGTAGCCC  1911
                :||| :::                                   |||
Db   416  rArgSerVal---------------------Pro  416
Qy  1912  TTCCAACAC---------CTGACAGAATGAGACCAAACGCAG--------CCAGCCAGA  1953
                |||:::     ||||||                        ::::::
Db   417  PheArgAspAspSerAlaMetThrMet------GlnThrGlnPheSerAlaProAlaGln  434
Qy  1954  TCG------GGAGCAAACCAAAGACCATCTGAGGAATGAGAGATCTGCGGAGGCGGCCAG  2007
                :::      ||||||:::                    ::::::
Db   435  AsnAsnProPheGlyLeuGlyGlnIleTrpAsnProValAlaGlnIleProAlaAlaAlaGln  454
Qy  2008  GGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCGGAGGGGCGGGGAAGGTCGCCAG  2067
                |||||                  :: |||
Db   455  ---------------AsnProTyrGlnValGlyGlyValGlyTrpGlnGlnAsnSerVal  467
```

```
Db    455 yProProValProHisGlyProHisGlyVal-ProGlyProHisGlyProProGlyProP 475
              |||       |||||||
Qy   1878 CAGGCACCAGCAAACAACAACGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCA 1937

Db    475 roGly-ProGly------------------------------ThrPro-----MetGlyPro 483
              |||  |||
Qy   1938 AACGGACCAGCCAGATCGGGAGCAACCAAAGACCAAATGACCATCTGAGGAATGAGAGTCTCGGG 1997

Db    484 TyrAsnProAlaProTyrAsnProGlyPro------------------------- 493
              :::
Qy   1998 AGGGGGCCAGGGACTCTGCCGAGG---CCCTGAGAACCCCAGGGGCCGAGGAGGGGCGGG 2054

Db    494 ------ProGlyProAlaProHisGlyProProAlaProTyrAlaProGlnGlyTrpGly 511
                    |||||:::
Qy   2055 GAAGGTCAGCCAGGTTTGCCAGAA------CCACCGAGCCCCGCC 2093

Db    512 AsnAlaTyrProHisTrpGlnGlnGlnAlaProProAla 526
```

RESULT 10
S58529
alpha-complex protein 1 - human
N;Alternate names: nucleic acid-binding protein; protein PCBP-1
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C;Accession: S58523; S65678; S43489; S41378; S42472
R;Kiledjian, M.; Wang, X.; Liebhaber, S.A.
EMBO J. 14, 435-4364, 1995
A;Title: Identification of two KH domain proteins in the alpha-globin mRNP stability com
A;Reference number: S58523; MUID:9616208; PMID:7556077
A;Accession: S58529
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-356 <KI1>
A;Accession: S58523
A;Status: preliminary
A;Molecule type: protein
A;Residues: 125-139;231-265;315-323 <KI2>
R;Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A;Title: Characterisation of two major cellular poly(rC)-binding human proteins, each co
A;Reference number: S65678; MUID:95331278; PMID:7607214
A;Accession: S65678
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-204,'V',206-356 <LEF>
A;Cross-references: EMBL:X78137; NID:g460770; PIDN:CAA55016.1; PID:g460771
A;Experimental source: AMA cells (transformed human amnion cells)
A;Note: submitted to the EMBL Data Library, March 1994
R;Aasheim, H.C.; Loukianova, T.; Deggerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994
A;Title: Tissue specific expression and cDNA structure of a human transcript encoding a
A;Reference number: S43489; MUID:94203810; PMID:8152927
A;Accession: S43489
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-298,'H' <AAS>
A;Cross-references: EMBL:Z29505; NID:g444020; PIDN:CAA82631.1; PID:g444021
C;Genetics:
A;Gene: GDB:HNRNPX
A;Cross-references: GDB:344947
C;Keywords: RNA binding

Alignment Scores:
Pred. No.:
Score: 238.50   Length: 356
Percent Similarity: 41.78%   Matches: 89
Best Local Similarity: 23.99%   Conservative: 66
Query Match: 3.90%   Mismatches: 137
DB: 2   Indels: 79
                                           Gaps: 13
Pred. No.:   2.06e-08

US-09-270-437d-6 (1-3412) x S58529 (1-356)
```

```
Qy    862 ---------GAGGCGGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC 911
                     |||               :::    :::|||
Db    175 oGlyPheHisHisGlyAspGlyProGlyAsnAlaValGln-------GluIleMetIl 192

Qy    912 ACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTGAAGAAAATTGA 971
              :::||||||| ||||:::||||||
Db    192 eProAlaSerLysAlaGlyLeuVaIIleGlyLysGlyGluThrIleLysGlnLeuGl 212

Qy    972 ACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATCAACCC 1031
              |||::::  |||:::  ::: ||||||
Db    212 nGluArgAlaGlyVaILysMetVaIMet------IleGlnAsp----------GlyPr 226

Qy   1032 GGAAAGAAACCATCACTGTGAAGGGACCAGTTGAGCCTGTGCCAGTGCTGAGATAGAGAT 1091
              |||     |||
Db    226 oGlnAsnThr---------------------GlyAlaAsp--------- 232

Qy   1092 TATGAGAAGCTGCGTGAGGGCCTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAA 1151
              ||||||     :::
Db    233 ---LysProLeuArg------IleThrGlyAspProTyrLysValGlnGlnAlaLysGl 249

Qy   1152 TCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCT 1211
              :::::                                           ||
Db    249 uMetVal------------------------------------------LeuGluLeuI 255

Qy   1212 ATCTCCACCAGCACAGGCCCCGGGAGCTCCCCCGCCTGCCCCCCTACCACCCCTTCACTAC 1271
              ||| |||||||       |||
Db    255 eArgAspGlnGlyPheArgGluVal------------------------------ 264

Qy   1272 CCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCGTTCCCGCATCA 1331
              :::::::::|||
Db    265 ------------ArgAsnGluTyrGly--------- 269

Qy   1332 TCACTCTTATCCAGAGCAGGAGATTGTGAATCCTCTTCATCCCAACCCAGGCTGTGGGCGC 1391
              ||| |||
Db    269 -SerArgIleGlyGlyAsnGluGlyIleAspGlyIleProArgPheAlaValGlyIl 289

Qy   1392 CATCATCGGAGAAGAGGGGGCCACACATCAAACAGTGGCGGAGATTCGCCGGAGGCCTCAT 1451
              ::||||||||:::::::      ||||
Db    289 eVaIIleGlyAspGlyAsnGlyValGluMetIleLeuLysLysGlnAsnAspAlaGlyValArgIl 309

Qy   1452 CAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGGAAAGGATGGTCATCATCACCGGGCC 1511
              ||| :::||
Db    309 eGlnPheLysProAspAspGly---ThrThrProGluArgIleAlaGlnIleThrGlyPr 328

Qy   1512 ACCGGAA---GCCCAGTTCAAGGCCCAG 1536
              ||||||
Db    328 oProAspArgCysGlnHisAlaAlaAlaGluIleIleThrAspLeuLeuArgSerValGlnAl 348

Qy   1537 ------GGACGATCTTGGGAAA 1554
                    |||:::
Db    348 aGlyAsnProGlyProGlyProGlyArgGlyArgGlyArgGlyGlnGlyAsnThr 368

Qy   1555 ------CTGAAAGAGGAGAAACTTCTTTAACCCCAAAGAAGAGAGT 1592
              :::
Db    368 pAsnMetGlyProGlyProGlyLeuGlnLeuGluProAsnPheIle----------- 382

Qy   1593 GAAGCTGGAAGCCGCATATCAGAGTGCCCCTTCCACAGTCGCCGGGTGATTGGCAAAGG 1652
              |||||:::        ||||||||||
Db    383 ------------ValProThrGlyLysThrGlyLeuIleGlyLeuIleGlyLysGl 395

Qy   1653 TGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGA 1712
              :::|||||| :::  ||| :::::
Db    395 yGlyLysGluThrIleLysSerIleSerGlnSerGlyAlaArgIleGlnLeuGlnArgAs 415

Qy   1713 CCAAACGCCAGATGAAAATGAGGAAGTG---ATCGTCAGAATTATCGGGCAC---------- 1761
              ::||| |||      |||  :::
Db    415 nProProAsnAlaAspProAsnMetLysPheThrIleArgGlyThrProGlnIl 435

Qy   1762 ---TTCTTTGCTAGCCAGACTGCCACAGTCGCAAGATCAGCAGGGAAATTGTACACAGGTGAA 1817
              :::|||        ||||||
Db    435 nIleAspTyrAlaArgGlnIleLeuGluGlyIleGluLeuIleSerIleGlyGlyProValAsnProLeuGl 455

Qy   1818 GCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCCAGCGCAAGTGAGGCTCCCA 1877
```

Db ||| :::||| |||
        340 GlnTyrLeuIleAsnAlaArgLeuSerSerGlu 350

RESULT 11
T04533
hypothetical protein F28J12.30 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04533
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04533
A;Molecule type: DNA
A;Residues: 1-846 <BEV>
A;Cross-references: EMBL:AL021710
A;Experimental source: cultivar Columbia; BAC clone F28J12
C;Genetics:
A;Introns: 94/3; 140/3; 161/3; 191/3; 232/3; 262/1; 287/1; 293/3; 300/3; 503/3; 675/3;
A;Map position: 4
A;Note: F28J12.30

Alignment Scores:
Pred. No.:   Length: 846
Score: 3.59e-08   Matches: 122
Percent Similarity: 236.50   Conservative: 92
Best Local Similarity: 35.14%   Mismatches: 229
Query Match: 20.03%   Indels: 166
DB: 3.86%   Gaps: 19
                     2

US-09-270-437D-6 (1-3412) x T04533 (1-846)

Qy   370 CTTTTGGCTCAAATATGGACAGTGGAGAATGTGGAACAAGTCAACACA------------------ 417
         ||||||||||| ::: |||||
Db   268 LeuLeuAspSerTyrArgGly------------HisThrIleGlyValAsnThrAlaThrPheThr 284

Qy   418 ------------------------------------GACACAGAAACCGCCGTTGTCAACGTC 444
                                            :::|||||| :::|||
Db   285 ArgLysGlySerGlyMetSerSerGlyLeuIleGluThrAspSerGluValLeuAsnSer 304

Qy   445 ACATATGCAACAAGAGAAGGAGCAAAATAGCCATGGGCATCAAGCGGGCATCCAGTT 504
         ||||||| ||||||
Db   305 ValTyrLeuThrMetValGluArgLys------------------------------ 313

Qy   505 GGAGAACTACTCCCTTCAAGATTCCTACATCCCGGGATGAAGAGGTGAGTCCCCCTTCGCCC 564
         |||||||||||| ||||||
Db   314 --Lys 314

Qy   565 CCTCAGCGAGCCCAGCCGTGGGGACCACTCTTCCCGGGACCAAGGCCACGCCCCTGGGGGC 624
         :::::|||||||||||| :::
Db   315 ArgLysGlnIleGlnArgAsnAsnSerGluSerAsnArg--------------------- 327

Qy   625 ACTTCTCAGGCCACCAGACAGATTGATTTCCCG---------------------------CTG 657
         ||||||||||| :::|||
Db   328 ---AsnGlnLysArgArgSerHisAspLysIleAsnArgGluLeuValTyr 346

Qy   658 CGGATCCCTGGTCCCCCAGTTGTTGGTGCCCATCATCGGGAAAGGAGGGCTTGACCATA 717
         ||||||||||| ||||||
Db   347 ArgIleLeuCysProIleAspValValGlyValIleGlyValLysLysValIle 366

Qy   718 AAGAACATCACTAAGCACAGCACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGA 777
         ||||||||||| :::
Db   367 AsnAlaIleArgHisAsnThrLysAlaLysIleLysValPhe---AspGlnLeuHisGly 385

Qy   778 GCTGCAGAGAAGCCTGTCACCATCCATGCCAAGGC----CCTGCCGAAGGC 837
         ||||||||||| ::::
Db   386 CysSerGlnArgValIleThrIleTyrCysSerValLysGluGlyGlnGlu--------- 403

Qy   838 ATGATTCTTGAAAATCATGCAGAAAGAGGCAGAATGAGACCCAAGACTAGCCCAAGAGATTCCT 897
         ||||||||||| :::
Db   404 ------IleGlyPheThrLysSerGluAsnGluProLeuCysCysAlaGlnAlaAspAlaLeu 421

Qy   898 CTGAAAATCTTGGCACACAATGGCTTGGTT-------------------------------- 927

Qy   619 GGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGTGCGGATCCTGGTCCCCACCCAG 678
         :::|||| ::|||
Db   4 GlyValThrGluSerGlyLeuAsnValThrLeuThrIleArgLeuMetHisGlyLys 23

Qy   679 TTTGGTGGTGCCATCATCGGGAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGCAGACC 738
         |||||| :::|||||||||||| :::|||
Db   24 GluValGlySerIleIleGlyLysGlyGlySerValLysArgIleArgGluGluSer 43

Qy   739 CAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCCAGAGAAGCCTGTCACC 798
         ::|||||||||:::| ::|||
Db   44 GlyAlaAspGlyAsnIle------------SerGluGlyAsnCysProGluArgIleIleThr 60

Qy   799 ATCCATGCCACCCCAGAGGGGACTTCTGAAGCCATGGCGCATGATTCTTGAAAATCATGCAG 858
         :::| ::|||
Db   61 LeuThrGlyProThrAsnAlaIlePheLysAlaPheAlaMetIleIleAspLysLysLeuGlu 80

Qy   859 AAAGAGGCAGATGAGACCAAACTA--------------GCCGAAGAGATTCCTCTG 900
         ::::: :::   :::||| :::|||
Db   81 GluAspIleAsnSerSerMetThrAsnSerThrAlaAlaSerArgProProValThrLeu 100

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTG 960
         |||||| :::|||
Db   101 ArgLeuValProAlaThrGlnCysGlySerLeuIleGlyLysLysGlyGlyCysLysIle 120

Qy   961 AAGAAAATTGAACATGAAACAAGGACCAAGATAACAATCCATCTTTGCAGGATTGAGC 1020
         |||||| |||||| ::|||
Db   121 LysGluIleArgGluThrGlyAlaGlnIleValAlaGly------AspMetLeu 138

Qy 1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTCTGCCAGTGCT 1080
         ::: |||||| :::||||||
Db   139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158

Qy 1081 GAGATAGAGATT------------ATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATG 1128
         ||||||||||||
Db   159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178

Qy 1129 CTGGCTGTTAACCAACAAGCCAATCTGATCCAGGTTGAACCTCAGCGCACTTGGCATC 1188
         :::   :::||||||
Db   179 MetThrIleProTyrGlnPro--------MetProAlaSerSerProValIle 193

Qy 1189 TTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCGCCCCGGGAGCTCCCCGCT 1248
         |||||| ||||||
Db   194 CysAlaGlyLysGlyAlaAspArgCysSerSerAspAlaAlaGly--------- 206

Qy 1249 GCCCCCTACCACCCCTTCACTACCCAC--------------TCCGGATACTTC 1287
         ::: |||||||||||| :::
Db   207 -------------TyrProHisAlaThrHisAspLeuGluGlyProProLeuAspAlaTyr-- 222

Qy 1288 TCCAGCCTGTACCCCCATCACCAGTTTGGCCCG--------------------------- 1320
         :::::||||| :::
Db   223 --SerIleGlnGlyIleHisThrIleSerProLeuAspLeuAlaLysLeuAsnGlnVal 241

Qy 1321 --------------------------TTCCCGCATCATCAC---------------------- 1335
                                   |||||||||||
Db   242 AlaArgGlnSerHisPheAlaMetMetHisGlyGlyThrGlyPheAlaGlyIleAspAsp 261

Qy 1336 --TCCTATCCAGAG----------------------------------CAGGAGATT 1356
         |||||||||| ||||||
Db   262 SerSerSerProGluValLysGlyTyrTrpTrpAlaSerLeuAspAlaSerThrGlnThrThr 281

Qy 1357 GTGAATCTCTTCATCCCAACCCAGGCTGTGCGGGCCATCATCGGGAAGAGGGGGCACAC 1416
         :::::|||||| |||||||:::
Db   282 HisGluLeuThrIleProAsnAsnLeuIleGlyCysIleLeuGlyArgGlnGlyAlaAsn 301

Qy 1417 ATCAAACAGCTGGCGAGATTCGCCGGAGCCTTCATCAGATTGCC---CCTGCCGAAGGC 1473
         :::   ::||||| ||||||
Db   302 IleAsnGlnIleArgGlnMetSerGlyAlaGlnIleIleLysValAlaAsnProValGluGly 321

Qy 1474 CCAGAGCTCAGCGAAGGAAGATGGTCATCATCACCGGGCCACCCAGTTCAAGGCC 1533
         :::|||||| ::|||||||
Db   322 Ser------SerGlyArgGlnValThrIleThrGlyThrProAlaAlaSerLeuAla 339

Qy 1534 CAGGGACGGATCTTTGGGAAACTGAAAGAGGAA 1566

```
Db   768  ------------------------SerSerValArgAspAlaLeuIleGlnIle-ValLeuAr  780
                                                                       :::  |||
Qy  1825  GAGCAGAAATACCCTCAGGAGTCGCCTCACAGCGCAGCAAGTGAGCGTCCCACAGGCAC        1884
          |||                                                               |||
Db   780  gLeuArgGluAspValLeuGlyAspLysAspSerValAlaThrArgLysProProAla--       799
                                                                         :::
Qy  1885  CAGGCAAAACAACGATGAATGTAGC                                         1909
          |||||||||
Db   800  ------ArgThrAspAsnCysSer                                          805
```

RESULT 12
S46109
hnRNP complex protein homolog YBR233w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR1531
C;Species: Saccharomyces cerevisiae
C;Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002
C;Accession: S46109
R;Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;
  submitted to the Protein Sequence Database, August 1994
A;Reference number: S45782
A;Accession: S46109
A;Molecule type: DNA
A;Residues: 1-413 <DUB>
A;Cross-references: EMBL:Z36101; NID:g536631; PIDN:CAA85196.1; PID:g536633; MIPS:YBR233w
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:PBP2
A;Cross-references: SGD:S0000437
A;Map position: 2R
C;Keywords: nucleus

Alignment Scores:
Pred. No.:
Score: 3.73e-08   Length: 413
Percent Similarity: 235.00   Matches: 102
Best Local Similarity: 40.20%   Conservative: 56
Query Match: 25.95%   Mismatches: 151
DB: 3.84%   Indels: 84
                          2   Gaps: 16

US-09-270-437D-6 (1-3412) x S46109 (1-413)

```
Qy   523  ATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCCAGCGT   582
          ||::|                          :::       ::::|
Db    10  IleThrThrProThrThrValLeuValSerProAsnThrLeuLysArg--LysLys      28

Qy   583  GGGGACCACTCTTCCCGGAGCAA                                        606
          ||||
Db    29  GlyGluAspThrSerGluGlnLeuGluAlaGluLysArgValAlaLeuLysLysAsp      48

Qy   607  --GGCCACGCCCCTGGGGGCCACTTCTCCAGGCACAGATT-----GATTCCCG         654
             ::|
Db    49  AlaAspSerHisSerAspAsnAspHisAspSerProAspAsnValProSerValHis      68

Qy   655  CTGCGGATCCTGGTCCCCACCCAGTTGTTGTGGCCATCGGAAAGGAGGGCCTTGACC     714
          |||                                      ::|||      :::|||
Db    69  LeuArgMetLeuCysLeuValLysHisAlaSerLeuIleValGlyHisLysLysGlyAlaThr 88

Qy   715  ATAAAGAACATCACTAGCACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCT       774
          ||::|||                  :::
Db    89  IleSerArgIleLysSerGluThrSerAlaArgIleAsnIleSerAsnIleArg---     107

Qy   775  GGAGCTGCAGAGAAGCCTGTCACCATCGCCACCCCAGAGGGACTTCTGAAGCATGC       834
          |||                    ::|
Db   108  GlyValProGluArgIleValTyrValTyrArgGlyThrCysAspAspValAlaTyr      127

Qy   835  CGCATGAATTCTTGAAATCATGCGAGAAGAGAGGCAGATGAGACCAAACTAGCCGAA-----  888
          |||                      :::
Db   128  GlyMetIleValArgAlaLeuLeuGluGluHisGlyValAsnGlyGluAspIle         147

Qy   889  GAGAGTTCCTCTGAAAATCTTGCACACCAATGGCTTTGGTTGGAAAAAAGAA           948
          |||
Db   148  GluIleuSerIleuAsnIleuLeuIleuProHisHisLeuMetGlyCysIleIleIleGlyLysArg 167
```

```
Alignment Scores:
Pred. No.:                  1.62e-07        Length:        479
Score:                      226.00          Matches:        97
Percent Similarity:         39.18%          Conservative:   95
Best Local Similarity:      19.80%          Mismatches:    148
Query Match:                3.69%           Indels:        150
                                            Gaps:           18
                            DB:

US-09-270-437D-6 (1-3412) x C86275 (1-479)

Qy      583 GGGGACCACTCTTCCCGGGAGCAAGGCACGCCCCTGGGGGCACTTCCAGGCCAGACAG  642
                |||         |||        ::         ||      ||   |||     |||
Db       10 GlyLysArgSerAsnLeuGlnSerGluPheThrGlyAsnGlyGlySerLysArgArgAsn  29

Qy      643 ATT------------------------------------------GATTTCCGTGCGATCCTG  666
                |||                                          ||   |||   |||
Db       30 LeuHisAspGluThrAspGlnAsnValIleAlaSerGluAspThrValTyrArgTyrLeu  49

Qy      667 GTCCCCACCAGTTTGTTGTGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGACAATC  726
                |||      :::      |||::||   ||::::::::||::|||||     |||
Db       50 CysProValLysLysThrGlySerIleIleIleGlyLysGlyGlyIleLeuLysGlnIle  69

Qy      727 ACTAAGCAGACCAGGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGAGCTGCAGAG  786
                :::||||||                 :::        ||
Db       70 ArgSerGluThrLysSerAsnMetArgIleAsn---GluAlaLeuProGlyLysCysGluGlu  88

Qy      787 AAGCCTGTCACCATCCATGCCACCCCCAGAGGGACTTCT------------------  825
                :::|||||||:::|||::||         |||
Db       89 ArgValValThrMetTyrSerThrAsnGluLeuAsnHisPheGlyAspAspGlyGlu  108

Qy      826 --------GAAGCATGCCCATGATTCTGAAATCATGCAGAAAGAGGCA  867
                        ||||         ::|||
Db      109 LeuValCysProAlaLeuAspAlaLeuPheLysValHisAspMetValAlaAspAla  128

Qy      868 ---------------GATGAGACCAAACTAGCCGAAGAG------ATTCCTCTG  900
                               ||||         :::||
Db      129 AspGlnAspAspGlyThrAspAspAsnAspLeuGlyGluLysGlnThrValThrVal  148

Qy      901 AAAATCTTGGCACCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960
                ::::::|||               :::||||||||||||
Db      149 ArgMetLeuValProSerAspGlnIleGlyCysValIleGlyValIleGlyGlnValIle  168

Qy      961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020
                :::  :::      ||          :::::
Db      169 GlnAsnLeuArgAsnAspThrArgAsnAlaGlnIleArgVal----------------  181

Qy     1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCC------  1074
                                            :::    ||||||
Db      182 ----------------------------------------IleLysAspHisLeuProAlaCysAlaLeuThr  192

Qy     1075 ----------------------------------------------AGTGCTGAG  1083
                                                              ||
Db      193 LeuSerHisAspGluLeuLeuLeuIleLeuGlyGluProLeuValValArgGluAlaLeu  212

Qy     1084 ATAGAGATTATGAGAAGCTGCGTGAG---------GCCTTTGAAAATGATATGCTGGCT  1134
                :::                      :::       ||||
Db      213 TyrGlnValAlaSerLeuLeuHisAspAsnProSerArgPheGlnHisLeuLeuLeuSer  232

Qy     1135 GTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTGCCCACTTGGCATCTTTTCA  1194
                |||   :::            ||||||||         :::   ||
Db      233 SerSerSerSerMetHisGlnProGlyAlaMetLeuMetSerAlaAlaLeuThrSer  252

Qy     1195 ACAGGACTGTCCGTCATCTCCACCAGCAGGCCCCCGGAGCTCCCCCGCTGCCCCC  1254
                :::
Db      253 Ser-----------------------------------------------  253

Qy     1255 TACCACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCTGGCATCACCACCAGTT  1314
                :::
Db      253 ------------------------------------------------------  253

Qy     1315 GGGCCGGTTCCGCCGCCGTCATCACTCCTTAT--------------  1353
                   :::                          |||
Db      254 HisArgAsnTyrArgValAlaValArgArgAspIleAlaAspAlaArgGlyGlu  268
```

```
Qy      949 GGCAGAAATTGAAGAGAAAATTGAACTGAACAGGGACCAAGATAACAATCTCATCTTTG  1008
                |||       |||::::||||||               :::::
Db      168 GlySerArgLeuArgGluIleGluIleAspLeuSerAlaAlaLysLeuPheAlaSerProAsn  187

Qy     1009 CAGGATTTGAGCATACAACACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCC  1068
                |||             :::::|||   :::|||   ::|||       :::|||
Db      188 Gln-----LeuLeuLeuSerAsnAspArgIleLeuThrLeuThrIleAsnGlyValProAspAla  205

Qy     1069 TGTGCCAGCTGCTGAGATAGAGATTATGAGAAGCTG---------------  1104
                |||
Db      206 IleHisIleAlaThrPheTyrIleSerGlnThrLeuLeuAsnPheGlnMetGluSerPro  225

Qy     1105 ----------------CGTGAGGCCTTGAA------------AATGATATGCTGGCT  1134
                                :::                        |||
Db      226 GlnLysAsnValLysArgSerIleTyrTyrGlnProThrGlnProPheAsnSerValLeuIle  245

Qy     1135 GTTAACCAACAAGCCAATCTGATC------------------------------  1158
                |||    :::
Db      246 AspHisSerGlnProAsnThrIlePheHisGlnIleArgAsnHisGlnTyrHisProSerAsp  265

Qy     1159 ---------CCAGGGTTGAACCTGCCACTTGGCATCTTTCAACAGGA  1200
                         |||||||         ||     |||
Db      266 LysLeuLeuSerTyrLysProAsnLysAsnLeuPro-------IleSerSerThrLeu  282

Qy     1201 CTGTCCGTGCTATCTCCACCAGCAGGCCCCCGGAGCCTGCCCCCTACCAC  1260
                |||:::||||                     :::       ||
Db      283 LeuSerMetAlaThrProGlnTyrThrThrAlaSerValAlaAsnAlaThrAlaPheGln  302

Qy     1261 CCC----------TTCACTACCCCGGATACTTCTCCAGCCTGTACCCC---------  1302
                |||          ||||||||         ||||||||
Db      303 ProAsnPheValIleProAsnValThrValLeuAspGlyProValIleSer----  319

Qy     1303 CATCACCAGTTGGCCCGTTCCGCCGCCATCAC--------TCTTATCCAGAGCAGGAG  1353
                |||                             :::||||||||||
Db      320 ----------------ProAlaProGlyAsnHisLeuLeuMetAsnPheValGlnGlnGlu  334

Qy     1354 ATTGTGAATCTTCTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGAAAGAAGGGGGCA  1413
                |||                          :::||||||||||||||
Db      335 Ile--------PheIleAspGluLysProGlyLysAsnValIleIleGlyAsnGlyGlyLys  351

Qy     1414 CACATCAAACAGCTGCGAGATCCCGAGCCTCTATCAAGATTGCC---CCTGCGGAA  1470
                |||   :::                      ||||
Db      352 HisIleAsnSerValLysGlySerThrGlyCysSerIleIleIleGluGlnAspProValGlu  371

Qy     1471 GGGCCCAGACGTCAGCGAAAGATGGTCATCATCACCGGG  1509
                |||||||||         :::|||   |||
Db      372 GlySer-------SerGluArgArgLeuThrIleArgGly  382
```

RESULT 13

C86275

hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: C86275
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:1130712
A;Accession: C86275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <STO>
C;Genetics:
A;Map position: 1

A;Cross-references: GB:AE005172; NID:g5080792; PIDN:AAD39302.1; GSPDB:GN00141

```
Qy   682 GTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGACATCATCATAGACAGACCCAG 741
             ||||||||||||||||||||||||  :::||||||||  ||||||||
Db    61 AlaGlySerIleIleGlyLysGlyGlyLeuThrIleValGlnLeuGlnLysGlyLysThrGly 80

Qy   742 TCCCGGTGAGATATCGAAGAGAAAGAGAACCT------GGAGCTGCAGAG 786
             :::  |||||                                      ||||||
Db    81 AlaThrIleLysLeuSerLysLysSerLysAspPheTyrProGlyThrThrGlu 100

Qy   787 AAGCCTGTCACCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCCATGATTCTT 846
             ||||  |||||||  :::  |||||                      :::
Db   101 ArgValCysLeuIleGlnIleGlyThrValGluAlaLeuAlaLeuAsnAlaValHisGlyPheIleGluAla 120

Qy   847 GAA------------------------------------------ATCATGCAG 858
             |||
Db   121 GluLysIleArgGluMetProGlnAsnValAlaLysThrGluProValSerIleLeuGln 140

Qy   859 AAAGAG------GCAGATGAGACCAAACTAGCCGAGAGAGATTCCT------ 897
             :::          :::      :::    :::||||
Db   141 ProGlnThrThrValAsnProAspArgIleLysGlnThrLeuProSerSerProThrThr 160

Qy   898 ------------------------------CTGAAAATCTTG 909
                                              |||  |||
Db   161 ThrLysSerSerProSerAspProMetThrSerArgAlaAsnGlnValLysIleIle 180

Qy   910 GCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAAAATT 969
             |||||||||||||||||||||||||||:::||||||||  :::|||
Db   181 ValProAsnSerThrAlaGlyLeuIleLeuIleGlyLysGlyAlaThrValLysAlaVal 200

Qy   970 GAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATCACAC 1029
             :::||||||||     |||  :::  |||
Db   201 MetGluGlnSerGlyAlaTrpValGlnLeuSerGlnLysProAsp------GlyIleAsnLeu 219

Qy  1030 CCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGTGCCAGTGCTGAGATAGAG 1089
             |||
Db   220 GlnGluArgValValThrValSerGlyGluProGluGlnAsnArgLysAlaValGluLeu 239

Qy  1090 ATTATGAAGAAGCTGCTGAG------------------------------ 1110
             |||:::||||:::|||
Db   240 IleIleGlnLysLysIleGlnGluAspProGlnSerGlySerGlySerCysLeuAsnIleSerSerTyrAla 259

Qy  1110 ---------------------------------------------- 1110
Db   260 AsnValThrGlyProValAlaAsnSerAsnProThrGlySerProTyrAlaAsnThrAla 279

Qy  1110 ---------------------------------------------- 1110
Db   280 GluValLeuProThrAlaAlaAlaAlaAlaGlyLeuLeuGlyHisAlaAsnLeuAlaGly 299

Qy  1111 ------------------AATGATATGCTGGCTGTGTTAAC 1140
                                 ||  :::::||||||
Db   300 ValAlaAlaAlaPheProAlaValLeuSerGlyPheThrGlyAsnAspAlaValAlaIleThr 319

Qy  1141 CAACAGCCAATCTGATCCCA-----GGGTGAACCTCAGCGCACTTGGCATCTTTTCA 1194
             |||||||||||||||:::::  |||||
Db   320 SerAlaLeuAsnThrLeuAlaSerTyrGlyTyrAsnLeuAsnThrLeuGlyLeu------ 337

Qy  1195 ACAGGACTGTCCGTGCTATCTCCACAGCCAGGCCCCGCGAGCT-------- 1239
             ||||||||||||||||||
Db   338 ---GlyLeuSer-----------GlnAlaAlaAlaAlaThrGlyAlaLeuAlaLeuAlaAlaAlaAlaAla 352

Qy  1240 --------CCCCCGTGCCCCTACCACCCCTTCACTACCCAC---------- 1275
                       ||||  :::
Db   353 AlaSerAlaAsnProAlaAlaAlaAlaAlaAsnLeuLeuAlaThrArgAlaSerGluAla 372

Qy  1275 ---------------------------------------------- 1275
Db   373 SerAlaSerGlyThrAlaGlyThrAlaGlyThrPheAlaLeuGlySerLeuAla 392

Qy  1276 ----------------------TCCGGATACTTCTCCAGCCGTGTACCCCCATCACCAGTTTGGC 1317
                                    :::  |||
Db   393 AlaAlaThrAlaAlaAlaThrAsnGlyTyrPheGlyAlaAlaSerProLeuAlaAlaSerAla 412

Qy  1318 CCGTTCCGCATCATCACTCTTTATCCA---GAGCAGGAGATTGTGAATCTTCATCCCA 1374
             :::||||||||||  |||||||      :::
Db    41 ArgThrAsnThrGlyValAspGlyGluAspProGlnTyrPheLeuLysValLeuLeuIleLeuLysProSerTyrAla 60
```

```
Qy  1354 ATTGTGAATCTTCATC---CCAACCCAGGCGTCTGGGCGCCATCATCGGGAGAGAAGGGG 1410
             ||||                  |||  :::||||||||||||||
Db   269 PheCysValCysPheIleCysProAlaGlyLeuGluAsnValGlyValIleGlyValIleGlyGly 288

Qy  1411 GCACACATCAAACAGCTGGCCGAGATTCGCCGGAGCCTCTATCAAGATTGCCCCTGCGAA 1470
             |||  :::          |||||  :::||||||:::
Db   289 GlyPheIleAsnGlnIleLeuArgGlnGluThrProGlyAlaThrIleArgValAlaAsnThrSerGlu 308

Qy  1471 GGGCCAGACGTCAGCGAAAGGATGGTCATCATCACGGGCACCGGAA---------- 1518
             ||||||||||||||||:::      :::
Db   309 ThrAspAsp------AspAspCysIleIleGluPheIleSerSerLysSerGluPheTyrGluAsp 326

Qy  1519 -------------GCCCAGTCAAGGCCCCAGGGACGGATCTTTGGGAAACTG--- 1557
                          |||  :::          |||  |||
Db   327 GlnSerProAlaValAlaAsnAlaAlaIleAlaIleArgLeuGlnGlnAsnArgCysSerSerGlyValGly 346

Qy  1558 AAAGAGGAAACTTCTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGCGCCATATCAGAGTG 1617
             ||||  :::      |||  |||
Db   347 -----------------AspLeuAlaIleSerGluThrArgArgLeuLeuLeuVal 360

Qy  1618 CCCTTCCCACAGCTGCCGCGGGTGATTGGCCAAAGGTGGCAAGACCGTGAACGAACTGCAG 1677
             ||||||  :::||||||||||
Db   361 SerSerSerGlnIleGlyLeuCysLeuIleGlyLysGlyLysGlyIleGluSerGluMetArg 380

Qy  1678 AACTTAACCAGTGCAGAAGTC--ATCGTGCCTCGTGACCAAAGCCCAGAT------GAA 1728
             :::::||||  |||  :::||||||  :::::::
Db   381 SerValThrArgAlaAsnIleLeuArgIleLeuGlnLysGluLeuGlyIleLeuGluIleLeuSerGlnAlaArg 400

Qy  1729 AATGAGGAAGTGATCGTCAGAAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGC 1788
             ||||||||||  :::  |||||||
Db   401 GluAspGluGluMetValGlnIleIleThrGly------SerProAspAlaAlaMetLys 417

Qy  1789 AAGATCAGGGAAATTGTACACAGTGAAG 1818
             :::  ::::::|||
Db   418 AlaLeuThrGlnInValIleLeuLeuArgLeuArg 427
```

```
RESULT 14
I38489
onconeural ventral antigen-1 - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999
C;Accession: I38489
R;Buckanovich, R.J.; Posner, J.B.; Darnell, R.B.
Neuron 11, 657-672, 1993
A;Title: Nova, the Paraneoplastic Ri Antigen, is Homologous to an RNA-Binding Protein an
A;Reference number: I38489; MUID:94000830; PMID:8399153
A;Accession: I38489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-510 <RES>
A;Cross-references: EMBL:U04840; NID:g440877; PIDN:AAA16022.1; PID:g440878
C;Genetics:
A;Gene: Nova-1
```

```
Alignment Scores:
Pred. No.:                           510
I38489
Score:          1.65e-07       Length:      510
               226.00          Matches:     105
Percent Similarity: 34.29%      Conservative: 62
Best Local Similarity: 21.56%   Mismatches:  158
Query Match:    3.69%            Indels:      162
DB:             2               Gaps:        16
```

```
US-09-270-437D-6 (1-3412) x I38489 (1-510)
```

```
Qy   562 CCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCGGAGCAAGGCCACGCCCCTGGG 621
             ||||||      :::||                       |||
Db    22 ProProAspSerArgLysArgProLeuGlnAlaProGluProGluAlaGly--SerThrLys 40

Qy   622 GGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCCGGTCCCCACCCAGTT 681
             ||||:::                                       :::
Db    41 ArgThrAsnThrGlyValAspGlyGluAspProGlnTyrPheLeuLysValLeuLeuIleLeuLysProSerTyrAla 60
```

```
Db   413  IleLeuGlyThrGluLysSerThrAspGlySerLysAspValValGluIleAlaValPro  432
                                :::::::|||                  :::|||
Qy  1375  ACCCAGCTGTGGGCGGCCATCATCGGAAGAAGAGGGGCCACACAAACAGCTGCGAGA 1434
          |||||||||||              ||||||
Db   433  GluAsnLeuValGlyAlaIleLeuGlyLysGlyThrLeuValGluTyrGlnGlu  452
                                    |||      :::   :::
Qy  1435  TTCGCCGGAGCCTCATCAAGATTGCCCCTGCCGAAGGC-------CCAGACGTCAGC 1485
          |||||||||:::   |||::::          :::|||    |||
Db   453  LeuThrGlyAlaArgIleGlnIleSer---LysLysGlyGluPheValProGlyThrArg  471
                                                     |||
Qy  1486  GAAAAGGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCCAGGGACGGATC 1545
                                    :::      :::
Db   472  AsnArgLysValThrIleThrProAlaAlaThrGlnAlaAlaThrGlnAlaAlaGlnTyrLeuIle  491
                                                        |||
Qy  1546  TTTGGGAAACTGAAAGAGGAA 1566
          ::::::
Db   492  ThrGlnArgIleThrTyrGlu 498

RESULT 15
S42471
hnRNP protein E2 - human
N;Alternate names: nucleic acid-binding protein; protein PCBP-2
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: S65679; S42471
R;Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A;Title: Characterisation of two major cellular poly(rC)-binding human proteins, each co
A;Reference number: S65678; MUID:95331278; PMID:7607214
A;Accession: S65679
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-365 <LEF>
A;Cross-references: EMBL:X78136; NID:g460772; PIDN:CAA55015.1; PID:g460773
A;Experimental source: AMA cells (transformed human amnion cells)
A;Note: submitted to the EMBL Data Library, March 1994
C;Keywords: RNA binding

Alignment Scores:
Pred. No.:            1.63e-07       Length:            365
Score:               225.50         Matches:           89
Percent Similarity:  41.40%         Conservative:      65
Best Local Similarity: 23.92%       Mismatches:        153
Query Match:         3.68%          Indels:            65
DB:                  2              Gaps:              14

US-09-270-437D-6 (1-3412) x S42471 (1-365)

Qy   601  GAGCAAGGCAGCCACCCGCCCTGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGG  660
                :::|||||||||||  :::          ::   :::
Db     2  AspThrGlyValIleGluGlyLeuGlyLeuAsn----------ValThrLeuThrIleArg  17
          ||:||:|||
Qy   661  ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAAGGAGGGCTTGACCATAAAG  720
          :::||||::   :::  ||||||||::|||||||  :::::|||
Db    18  LeuLeuMetHisGlyValGlyLysGluValGlySerIleIleGlyLysGlySerValLys  37
                                                          |||
Qy   721  AACATCACTAAGCACGACCCAGTCCCGGGTAGATATCCAGGTGATATTCCATAGAAAAGAAGAACTCTGGAGCT  780
              ::        ||:::::::                       |||
Db    38  LysMetArgGluGluSerGlyAlaArgIleAsnIle-------SerGluGlyAsnCys  54
                                                |||
Qy   781  GCAGAGAAGCCGTCACCATCCATGCCAGCCCCAGAGGGACTTCTGAAGCATGCCGCATG  840
                ||:::::::        :::|||
Db    55  ProGluArgIleIleThrLeuAlaGlyProThrAsnAlaIlePheLysAlaAlaMet  74
          |||
Qy   841  ATTCTTGAAATCATGCAGAGAGGCAGATGAGACCAAACTA------------------:.  882
          :::  ::::::::::  :::      :::
Db    75  IleIleAspLysLeuGluAspIleSerSerMetThrAsnSerThrAlaAlaSer  94
                      |||
Qy   883  GCCGAAGAGATTCCTGAAAATCTTGGCACACAAATGCCTTGGTTGGAAGACTGATTGGA  942
          ||||::::::::::::::
Db    95  ArgProValThrLeuArgLeuValValProAlaSerGlnCysGlySerLeuIleGly  114
```

```
Qy   943  AAAGAAGGCAGAGAAATTTGAGAAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCA 1002
          ||||                              :::::|||        :::::|||
Db   115  LysGlyGlyCysLysIleLysIleGluThrGlyAlaGlnGlnValValLysLysAlaValValAla  134
                                            ||||||
Qy  1003  TCTTTGCAGGATTGAGCATATACAACCCGGAAAAGAAACCATCACTGTGAAGGGCACAGTT 1062
          |||:::      :::   :::   ||||||!!:::
Db   135  Gly------AspMetLeuProAsnSerThrGluArgIleAlaIleThrIleAlaGlyIlePro  152
                                                       |||
Qy  1063  GAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAGCTGCGTGAGCCTTTGAAAAT 1122
          :::::::                                           :::|||
Db   153  GlnSerIleIleGluCysValLysGlnIleCysValValMetGluThrLeuSerGln  172
                            :::    :::
Qy  1123  GAT------ATGCTGGCTGTAACCAACAAGCCAATCTGATCCCAGGGGTG 1167
          |||                                    :::  :::
Db   173  SerProProLysGlyValThrIleProTyrArgProLysProSerProValIle  192
                                         :::|||||||||
Qy  1168  AACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCT----CCACCAGCA 1224
          :::                            :::|||||||
Db   193  PheAlaGlyGlyGlnAspArgTyrSerThrGlySerAspSerAlaSerPheProHisThr  212
                                            |||
Qy  1225  GGGCCCCGC------------------------GGAGCTCCCCCCGCTGCCCCCTACCACCCC 1263
              |||
Db   213  ThrProSerMetCysLeuAsnProAspLeuGluGlyLeuProProLeuGluAlaTyr----- 230
                                                              CACCAG 1311
Qy  1264  TTCACTACCACTCCGGATACTTCTCCAGCCTGTACCCCAT------------------||||||
          |||     ||||||
Db   231  ------ThrIleGlnGlyGlnTyrAlaIleProGlnProAspLeuThrLysLeuHisGln  248
                                                                |||
Qy  1312  TTT------GGCCCGTTCCCGCCATCATCAC------------------------------ 1335
          |||
Db   249  LeuAlaMetGlnGlnSerHisPheProMetThrHisGlyAsnThrGlyPheSerGlyIle  268
                                                      |||||||
Qy  1336  ------TCTTATCCAGAGCAGGAG------------------------------ 1353
                :::|||||||
Db   269  GluSerSerSerProGluValLysGlyTyrTrpGlyLeuAspAlaSerAlaGlnThrThr  288
                                      :::||||||
Qy  1354  ATTGTGAATCTCTTCATCCCAACCCAGGTGTGGCGGGCCATCATCGGAAGAAGGGGCA 1413
          |||||||||:::        :::|||||||
Db   289  SerHisGluLeuThrIleProAsnAspLeuIleLeuGlyArgGlnGlyAla  308
                                :::|||
Qy  1414  CACATCAAAACAGCTGGCCAGATTCGCCGGAGCCTCATCAAGATTGCC---CCTGCGGAA 1470
          |||:::   :::      |||||||
Db   309  LysIleAsnGluIleArgGlnMetSerGlyAlaGlnIleLysLysIleAlaAsnProValGlu  328
                                                        |||
Qy  1471  GGGCCAGACGTCAGCGAAAGGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAG 1530
          :::|||||||||||   :::||||||
Db   329  GlySer------ThrAspArgArgGlnValThrIleThrGlySerAlaAlaSerIleSerLeu  346
                            |||||||
Qy  1531  GCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAA 1566
          :::|||
Db   347  AlaGlnTyrLeuIleAsnArgLeuSerSerGlu  358
```

Search completed: July 16, 2004, 10:33:42
Job time : 97.5 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:   July 16, 2004, 10:30:58 ; Search time 31.5 Seconds
                                    (without alignments)
                          11280.214 Million cell updates/sec

Title:            US-09-270-437D-6
Perfect score:    6121
Sequence:         1 ggcagcggaggaggcgagga........aacctgaaatgttattt 3412

Scoring table:    BLOSUM62
                  Xgapop 10.0 , Xgapext  0.5
                  Ygapop 10.0 , Ygapext  0.5
                  Fgapop  6.0 , Fgapext  7.0
                  Delop   6.0 , Delext   7.0

Searched:         141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :        SwissProt_42:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USTO_spool_p/US09270437/runat_16072004_113049_13315/app_query.fasta_1.3591
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LCOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437 @CGN_1_1_46 @runat 16072004 113049 13315 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID		Description
1	278	4.5	721	1	FUB2_RAT	Q99pf5	rattus norv
2	262.5	4.3	572	1	FUB3_HUMAN	Q96124	homo sapien
3	260.5	4.3	769	1	ZBP2_CHICK	Q8uvd9	gallus gall
4	251	4.1	707	1	FUB2_HUMAN	Q92945	homo sapien
5	245.5	4.0	339	1	PCB3_HUMAN	P57721	homo sapien
6	244.5	4.0	339	1	PCB3_MOUSE	P57722	mus musculu
7	244	4.0	651	1	FUB1_MOUSE	Q91wj8	mus musculu
8	242.5	4.0	643	1	FUB1_HUMAN	Q96ae4	homo sapien
9	238.5	3.9	356	1	PCB1_RABIT	O19048	oryctolagus
10	238.5	3.9	606	1	Y475_ARATH	P58223	arabidopsis
11	238	3.9	492	1	NCA2_HUMAN	Q9unw9	homo sapien
12	235.5	3.8	474	1	NOA1_RAT	Q80wa4	rattus norv
13	235	3.8	413	1	PBP2_YEAST	P38151	saccharomyc
14	234.5	3.8	356	1	PCB1_HUMAN	Q15365	homo sapien
15	231.5	3.8	403	1	PCB4_HUMAN	P57723	homo sapien
16	231.5	3.8	403	1	PCB4_MOUSE	P57724	mus musculu
17	226.5	3.7	493	1	NOA1_MOUSE	Q9jkn6	mus musculu
18	226	3.7	510	1	NOA1_HUMAN	P51513	homo sapien
19	225.5	3.7	365	1	PCB2_HUMAN	Q15366	homo sapien
20	225	3.7	362	1	PCB2_MOUSE	Q61990	mus musculu
21	203.5	3.3	1268	1	VGLN_HUMAN	Q00341	homo sapien
22	201.5	3.3	1270	1	VGLN_CHICK	P81021	gallus gall
23	187	3.1	606	1	TDRH_HUMAN	Q9y2w6	homo sapien
24	185.5	3.0	463	1	ROK_HUMAN	Q07244	homo sapien
25	185.5	3.0	463	1	ROK_RABIT	O19049	oryctolagus
26	184.5	3.0	629	1	PAB2_ARATH	P42731	arabidopsis
27	179	2.9	5262	1	MLL2_HUMAN	O14686	homo sapien
28	176.5	2.9	470	1	NR54_HUMAN	Q15233	homo sapien
29	176	2.9	1222	1	S160_YEAST	P06105	saccharomyc
30	174.5	2.9	817	1	VRP1_YEAST	P37370	saccharomyc
31	174.5	2.9	1781	1	AK12_HUMAN	Q02952	homo sapien
32	169	2.8	1943	1	PC15_MOUSE	Q99pj1	mus musculu
33	168.5	2.8	5147	1	PCLO_HUMAN	Q9y6v0	homo sapien
34	165	2.7	2161	1	SHK1_HUMAN	Q9y566	homo sapien
35	164	2.7	2167	1	SHK1_RAT	Q9wv48	rattus norv
36	164	2.7	5703	1	MU5B_HUMAN	Q9hc84	homo sapien
37	162	2.6	653	1	PABP_SCHPO	P31209	schizosacch
38	161	2.6	660	1	YHL1_EBV	P03181	epstein-bar
39	158	2.6	2842	1	APC_RAT	P70478	rattus norv
40	157	2.6	1687	1	VIT2_FUNHE	Q98893	fundulus he
41	156	2.5	1329	1	KF10_HUMAN	Q9p218	homo sapien
42	155.5	2.5	2517	1	NCR2_HUMAN	Q9y618	h nuclear r
43	155	2.5	381	1	YBD2_YEAST	P38199	saccharomyc
44	153.5	2.5	636	1	PAB1_HUMAN	P11940	homo sapien
45	153.5	2.5	636	1	PAB1_MOUSE	P29341	mus musculu

ALIGNMENTS

```
RESULT 1
FUB2_RAT
ID   FUB2_RAT      STANDARD;      PRT;   721 AA.
AC   Q99PF5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 2 (FUSE binding protein 2) (KH
DE   type splicing regulatory protein) (KSRP) (MAP2 RNA trans-acting
DE   protein 1) (MARTA1).
GN   FUBP2 OR KHSRP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 73-87; 89-108 AND 475-486, SUBCELLULAR
RP   LOCATION, AND FUNCTION.
RC   TISSUE=Brain;
RX   MEDLINE=22246918; PubMed=12358751;
RA   Rehbein M., Wege K., Buck F., Schweizer M., Richter D., Kindler S.;
RT   "Molecular characterization of MARTA1, a protein interacting with the
RT   dendritic targeting element of MAP2 mRNAs.";
RL   J. Neurochem. 82:1039-1046(2002).
CC   -!- FUNCTION: Part of a ternary complex that binds to the downstream
CC       control sequence (DCS) of the pre-mRNA. Mediates exon inclusion in
CC       transcripts that are subject to tissue-specific alternative
CC       splicing. May interact with single-stranded DNA from the far-
CC       upstream element (FUSE). May activate gene expression (By
CC       similarity). Binds to the dendritic targeting element and may play
CC       a role in mRNA trafficking.
CC   -!- SUBUNIT: Part of a ternary complex containing FUBP2, PTBP1, PTBP2
CC       and HNRPH1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC       the cytoplasm of neuronal cell bodies and dendrites.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   ----------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC
DR    EMBL; AF308818; AAG59811.1; -.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH_type_1.
DR    Pfam; PF00013; KH; 4.
DR    SMART; SM00322; KH; 4.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
KW    Transport; mRNA transport; mRNA processing; mRNA splicing;
KW    Transcription regulation; Trans-acting factor; Nuclear protein;
KW    DNA-binding; RNA-binding; Repeat.
FT    DOMAIN       145    209    KH 1.
FT    DOMAIN       234    300    KH 2.
FT    DOMAIN       323    387    KH 3.
FT    DOMAIN       425    492    KH 4.
FT    REPEAT       572    685    4 X 12 AA IMPERFECT REPEATS.
FT    REPEAT       572    583    1.
FT    REPEAT       618    629    2.
FT    REPEAT       644    655    3.
FT    REPEAT       674    685    4.
FT    DOMAIN         7     68    GLY/PRO-RICH.
FT    DOMAIN        69    498    GLY-RICH.
FT    DOMAIN       499    613    ALA/GLY/PRO-RICH.
SQ    SEQUENCE     721 AA;  74226 MW;  482C7A765C60EE4A CRC64;

Alignment Scores:
Pred. No.:
Score:                 4.46e-10           Length:           721
Percent Similarity:    278.00             Matches:          178
Best Local Similarity: 33.89%             Conservative:     87
Query Match:           22.76%             Mismatches:       251
DB:                    4.54%              Indels:           266
                       1                  Gaps:             37

US-09-270-437D-6 (1-3412) x FUB2_RAT (1-721)

Qy      1 GGCAGCGGAGGAGGAGGCGGAGGAGCGCGGGCCGGGGGAGCCGGGGCTCTCGGGG        60
          |||::|||||||||||
Db     25 GlyAlaGlyGlyGlyGlyProProProGlyProProGlyAlaGlyAspArg-GlyGlyGlyGl   44

Qy     61 AAGAGACGGATGAGATGAACAAGCTTACATCGGACAAGCTTCCGCCCGCCCGTCACCGCCGAC  120
          ||||||||||
Db     44 yProGlyGly---------------------------------------------GlyGlyProGl  51

Qy    121 GACCTCCGGCAGCTCTTTGGGACGACGGAGAGCTGCCCCTGCCGGGACAGGTCCTGCTGAAG  180
          |||:||||||||
Db     51 yGlyGlyGlyAlaSer-GlyGlyGlyProSerGlnProProGlyGlyGlyGlyProGlyIleA   71

Qy    181 TCCGGCTACGGCCTTCGTGGACTACCCCGACCAGAACTGGGCCCATCCGCGCCATCGAGACC  240
          ||||:||||
Db     71 rgLys-AspAlaPheAlaAsp---------------------AlaValGlnArgAlaArgGln   84

Qy    241 CTCTCGGGTAAAGTG----------------------------------------------  255
          :::::||||||:::
Db     85 IleAlaAlaLysIleGlyGlyAspAlaAlaThrThrValAsnAsnAsnThrProAspPhe  104

Qy    256 --------GAATTGCATGGAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTA  306
                  :::::|||  ::||||||||||||
Db    105 GlyPheGlyGlyGlnLeuGlyAspGluArgGlyMetVal----------SerLysLysLeu  124

Qy    307 AGGAGCAGGAAAAATTCAGATTCGA--------------AACATCCCTCCACCTG  348
          ||||||||||||||
Db    125 AlaSerGlnGlnLysAspSerIleGlySerGlnLeuGlyProIleHisProProArgThr  144

Qy    349 --------------CAGTGGGAGGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTG---  393
                        :::::||||||||||||||:::::
Db    145 SerMetThrGluGlyTyrArgValProAspGlyMetVal---------GlyLeuIleIle  161

Qy    394 --------GAGAATGTGGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAAC  441
                  :::||||  ::::::||||||||
Db    162 GlyArgGlyGlyGluGlnIleIleAsnLysIleGlnGlnAspSer--------  175

Qy    442 GTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGACGAAGCGTAAGCGGGCCATCAG  501
          :::  :::  :::::|||
Db    176 --------------------------GlyCysLysValGlnIleSerProAspSerGlyGly---  187

Db    176
Qy    502 TTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGAGTGAGCTCCCCTTCG  561
          |||||||||||
Db    188 -----------------------------------LeuProGluArgSerLeuThrArgGly  197

Qy    562 CCCCCTCAGCGGCGAGCCCAGCGTGGG------------------GACCACTCTTCCCGGGAGCAA  606
          ||:::  |||:::
Db    198 AlaProGluSerValGlnLysAlaLysMetLeuAspAspSerIleValSerArgArgGlyArg  217

Qy    607 GGCCACGGCCCCTGGG----------------------------------------GGCAACTCTCAG  633
          ||||||
Db    218 GlyGlyProProGlyGlnPheHisAspAsnAlaAsnGlyGlyGlnAsnGlyThrValGln  237

Qy    634 GCCAGACAGATTGATTCCCGTGCGGATCCTGGTCCCACCCAGTTGTTGGTGCCATC  693
          |||:|||
Db    238 -----------------GluIleMetIleProAlaGlyLysAlaGlyLysLeuVal  249

Qy    694 ATCGGAAAGGAGGGGCTTGACCATAAAGAACATCACTAAGACGAGACCCAGTCCCGGTGATAGAT  753
          |||||||||  |||||||
Db    250 IleGlyLysGlyGlyGluThrIle-----------LysGlnLeuGlnGluArgAlaGly  265

Qy    754 ATCCATAGA---------------------------AAAGAACTCTGGAGCTCCAGAAGAAGCCTGTC  795
          :::
Db    266 ValLysMetIleLeuIleLeuGlnAspGlySerGlnAsnThrAsn--ValAspLysProLeu  284

Qy    796 ACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCCGCATGATTCTTGAAATCATG  855
          |||  :::|||:::::::||||||:::|||
Db    285 ArgIleIleGlyAspProTyrLysValGlnAlaCysGluMetValMetAspIleLeu  304

Qy    856 CAGAAGAGGCAGATGAG----------------------------ACCAAACTAGCC  885
          ::|||||||||
Db    305 --ArgGluArgAspGlnGlyGlyPheGlyAspArgAsnGluTyrGlySerArgValGly  323

Qy    886 GAAGAGATTCCTCTGAAAATCTGGCACACAATGGCTTGGTTGGTTGGAAGACTGATTGGAAAA  945
          ||||:::|||||||
Db    324 GlyGlyIleAspValProValProArgHisSer------ValGlyValIleGlyArg  341

Qy    946 GAAGGCAGAAATTGAAGAAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCT  1005
          :::||||||||||||  ::|||||||
Db    342 SerGlyGluMetIleLysIleGlnAsnAspAlaLysValArgIleGlnPheLys--  360

Qy   1006 TTGCAGATTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAG  1065
          ||||||  |||||
Db    361 ---GlnAspAspGlyThr---GlyProGluLysIleIleAlaHisIleMetGlyProProAsp  378

Qy   1066 GCCTGTGCCAGTGCTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGAT  1125
          ||||||||||||
Db    379 ArgCysGluHisValAlaAlaAlaArgIleIle------------------AsnAsp  389

Qy   1126 ATGCTGGCTGTGTAACCACAGCCAATCTGATCCCAGGGTTGAACCTGCAGCACTTGGC  1185
          :::::::
Db    390 LeuLeuGln----------------------------------------------  392

Qy   1186 ATCTTTTCAACAGGACTGTCCTGCTATCTCCACCAGCAGGCGCCCCGGCTCCCCCCCC  1245
          |||||||||
Db    393 -----------------SerLeuArgSerGlyTyrProProGlyProGlyAlaProGly  406

Qy   1246 GCTGCCCCCTCACCACCCTTCACTACCCACCTCCGATACTTCTCCAGCCTGTACCCCCAT  1305
          |||
Db    407 MetProPro---------------------------------------GlyLysArgGlyArgGlyArgGlyGln  418

Qy   1306 CACCAGTTGGCCGGTTCCCGCATCATCACTCTTATCCAGACAGAGATTGTGAATCTC  1365
          ::|||||||||
Db    419 GlyAsnTrpGlyPro---------------------ProGlyGlyGlu--MetThrPhe  430

Qy   1366 TTCATCCCAACCCAGCGTGTGGGCGCCATCATCGGGAAGAAGGGGCACACAAACAG  1425
          ::|||||||:::  ::::::||||||
Db    431 SerIleProThrHisLysCysGlyLeuValIleLeuGlyArgGlyGlyGluAsnValLysLysAla  450

Qy   1426 CTGGCGAGATTCGCCGGAGCCTCTATCAAGATTGCC-----------CCTGCGGAAGGCCCA  1476
          :::  :::  :::::|||
```

```
Db   451  IleAsnGlnGlnThrGlyAlaPheValGluIleSerArgGlnLeuProProAsnGly--  469

Qy  1477  GACGTCACGGAAAGGATGGTCATCATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAG  1536
          |||     :::::     :::||      |||     |||:::
Db   470  AspProAsnPheLysLeuPheValIleArgGlySerProGlnGlnIleAspHisAlaLys  489

Qy  1537  GGACGGATCTTTGGGAAACTGAAAGAGGAGAAAACTTCTTTAACCCCAAAGAAGAGTGAAG  1596
          |||
Db   490  GlnLeuIle-----------------------------------------GluGluLys  495

Qy  1597  CTGGAAGGCCATATCAGAGTGCCCTTCCACAGCTGGCGCGGGTGATTGGCAAGGTGGC  1656
          :::||     |||
Db   496  IleGluGly-------------ProLeuCysProValGlyPro------GlyProGlyGly  509

Qy  1657  AAGACCGTGAACGAACGACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAA  1716
                                                             |||
Db   510  -----------------------ProGlyProAla  513

Qy  1717  ACGCCAGATGAAAATGAGGAAGTGTCAGAATTATCGGGCACTCTTCTTGCTAGCCAG  1776
                                         :::|||||
Db   514  GlyPro-----------------MetGlyProPheHisProGlyPro  523

Qy  1777  ACTGCACAGCGCCAAGATCAGGAGAAATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATAC  1836
                                                   |||
Db   524  PheAsnGlnGly-----------------------  527

Qy  1837  CCTCAGGGAGTCGCCTCACAGCGGCAGCAAGTGAG------------------GCTCCCACAGG  1881
          |||     ::::|||     |||
Db   528  ProProGly-AlaProHisAlaGlyGlyProProHisGlnTyrProProGlnGl  547

Qy  1882  CACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACG  1941
          |
Db   547  yTrpGlyAsnThrTyrProGluTrpGlnPro-ProAlaProHis------AspProAsnL  565

Qy  1942  CAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGA--------ATGAGAAGTC  1992
          ||||||     :::     |||
Db   565  ysAlaAlaAlaAlaAlaThrAspProAsnAlaAlaAlaTrpAlaAlaTyrTyrSerHisTyrT  585

Qy  1993  TGCGGAGGCGGCCAGGGACTCTGCCGAGGGCCCTGAGAACCCCAGGGGGCCGAGGAGGGCG  2052
          :::|||||||     :::|||
Db   585  yrGlnGlnProProGlyProValPro--------------------GlyProAlaProAlaP  599

Qy  2053  GGGAAGGTCAGCCAGGTTGCCAGAACCACCGAGCCCGAGCCCGCCTCCGCCCCCCAG  2106
          ||||||     ||||||
Db   599  roAlaAlaProProAlaGlnGlyGlnProGln-ProProProThrGlyGln  616
```

RESULT 2
FUB3_HUMAN
ID   FUB3_HUMAN   STANDARD;   PRT;   572 AA.
AC   Q96I24; Q92946; Q93VB6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 3 (FUSE binding protein 3).
GN   FUB3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=97094955; PubMed=8940189;
RA   Davis-Smyth T., Duncan R.C., Zheng T., Michelotti G., Levens D.;
RT   "The far upstream element-binding proteins comprise an ancient family
RT   of single-strand DNA-binding transactivators.";
RL   J. Biol. Chem. 271:31679-31687(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Cervix, and Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May interact with single-stranded DNA from the far-
CC   upstream element (FUSE). May activate gene expression.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q96I24-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q96I24-2; Sequence=VSP_008323, VSP_008324;
CC   Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Detected in a number of cell lines.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC   frameshift in position 18.
CC   -!- CAUTION: Ref.2 (AAH01325) sequence differs from that shown due to
CC   a frameshift in position 527.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U69127; AAC50893.1; ALT_FRAME.
DR   EMBL; BC001325; AAH01325.1; ALT_FRAME.
DR   EMBL; BC007874; AAH07874.1; -.
DR   Genew; HGNC:4005; FUBP3.
DR   MIM; 603536; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; Repeat; Alternative splicing.
FT   DOMAIN   77   141   KH 1.
FT   DOMAIN   162   228   KH 2.
FT   DOMAIN   253   317   KH 3.
FT   DOMAIN   354   421   KH 4.
FT   VARSPLIC   1   64   MAELVQGQSAPVGMKAEGFVDALHRVRQIAAKIDSIPHLNN
FT   STPLVDPSVYGVGYGVQKRPLDDGV -> MPPI (in
FT   isoform 2).
FT   /FTId=VSP_008323.
FT   VARSPLIC   298   548   Missing (in isoform 2).
FT   /FTId=VSP_008324.
FT   CONFLICT   72   72   V -> D (IN REF. 1).
SQ   SEQUENCE   572 AA;   61640 MW;   F1BE22542BC197D CRC64;

Alignment Scores:

Pred. No.:		Length:	572
Score:	4.06e-09	Matches:	119
Percent Similarity:	262.50	Conservative:	82
Best Local Similarity:	37.43%	Mismatches:	204
Query Match:	22.16%	Indels:	133
	4.29%		

```
DB:                  1          Gaps:          17

US-09-270-437D-6 (1-3412) x FUB3_HUMAN (1-572)

Qy   496 CATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGCTCC 555
             ||||:|::::  |||||:::
Db    24 HisArgValArgGlnIleAlaIleAlaLysIleAspSerIlePro-----HisLeuAsnAsn 41

Qy   556 CCTTCGCCCCTCAGCGAGCCCAGCGTG---------GGGACCACTCCTTCC 597
             :::||||||| :::||||||||
Db    42 SerThr-ProLeuValAspProSerValTyrGlyTyrGlyValGlnLysArgProLeuAs 61

Qy   598 CGGGAGCAAGGCCACGCCCCTCGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCCGCTG 657
             ||||:|::        :::||||:::
Db    62 pAspGlyValGlyAsnGlnLeuGlyAlaLeuValHisGlnAsnArgThrValIleThrGlu-- 80

Qy   658 CGGATCCTGCTGCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATA 717
             ||||||:::
Db    81 GluPheLysValProAspLysMetValGlyPheIleIleGlyArgGlyGlyGluGlnIle 100

Qy   718 AAGAACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGA 777
             ||      :::|||:::||||||
Db   101 SerArgIleGlnIleGluSerGlyCysLysIleGlnIleile--AlaSerGluSerSerGly 119

Qy   778 GCTGCAGAGAAGCCTGTCACCATCCACCATGCCCAGAGGGGACTTCTGAAGCATGCCGC 837
             |||:::           :::|||||
Db   120 IleProGluProCysValLeuThrGlyThrProGluSerIleGluGlnAlaLysArg 139

Qy   838 ATGATTCTTGAAATCATGCAGAAA------------------GAGGCAGAT 870
             ::::|||:::          :::|||
Db   140 LeuLeuGlyGlnIleleValAspArgCysArgAsnGlyProGlyPheHisAsnAspIleAsp 159

Qy   871 GAGACCAAACTAGCCGAAGAGATTCCTCGAAAAATCTTGGCACACAAATGGCTTGGTTGGA 930
             ||||                    :::|||||
Db   160 SerAsnSerThrIleGlnGlu-----------IleLeuIleProAlaSerValGly 175

Qy   931 AGACTGATTGGAAAAGAAGGCAGAAATTGAACAAATTGAACAGGACCAAG 990
             |||:::||            :::|||||
Db   176 LeuValIleleGlyArgGlyGlyGluThrIleLysGlnLeuGlnLeuGlnLeuGlyValGlyTyr 195

Qy   991 ATAACAATCTCATCTTGCAGGATTGAGCATA--TACAACCCGGAAAGAACCATCACT 1047
             ::  ::    :::|||:::        ::::::
Db   196 MetValMet-----IleGlnAspGlyProThrGlyAlaAspLysProLeuArg 213

Qy  1048 GTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGTGCGT 1107
             |||||      :::          :::||
Db   214 IleThrGlyAspAlaPheLysValGlnIleAsnIleAlaArgGluMetValLeuGlyIleleIleArg 233

Qy  1108 GAGGGCCTTTGAAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAATCTCATCCCCAGGGTTG 1167
             |||   :::|||
Db   234 GluLysAspGlnIleAsnAspPhe--------- 240

Qy  1168 AACCTCCAGGCGCACTTGGCATCTTTTCAAGACGACTGTCCGTGTATCTCCACCAGCAGG 1227
             |||                         :::|||
Db   241 --------------------ArgGly 242

Qy  1228 CCCCGCGGAGCTCCCCCGCTGCCCCTTCACTCACCACTCCGGATACTTC 1287
             |||             :::          ::
Db   243 ValArgGly--------AspPheAsnSerArgMetGlyGlyGly 254

Qy  1288 TCCAGGCCTGTACCCCATCACCATCATCACCTTCCGCCATCATCACTCTTATCCCAGAG 1347
             ||
Db   255 Ser--------- 255

Qy  1348 CAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGGGGCCCATCATCGGGAAGAAG 1407
             :::::        :::|||:::::::||
Db   256 --------IleGluValIleSerValProArgPheAlaValGlyIleleValIleleGlyTyrArgAsn 272

Qy  1408 GGGGCACACATCAAACAGTCTGGCGAGATTCGCCGAGCCTCTATCAAGATTGCCCCTGCG 1467
             |||:::|||       :::|||
Db   273 GlyGluMetIleleLysLysGlnAspAlaArgIleGlnProHisLysProAsp 292

Qy  1468 GAAGGCCCAGACGTCAGCGAAGATGGTCATCATCACCGGGCCACCGGAAGCCAGTTC 1527
             :::||
Db   293 AspGlyIleSer---ProGluArgAlaAlaAlaGlnValMetGlyValProProAsp--ArgCys 310

Qy  1528 AAGGCCCAGGACGGACGGATCTTTGGGAAACTG--------AAAGAGGAAAAACTTCTTT 1575
             :::          |||  :::|
Db   311 GlnHisAlaAlaHisIleIleSerGluLeuThrLeuArgGlnGlnArgAspGlyPhe 330

Qy  1576 AACCCCAAAGAAGAAGTGAAG-------------------------- 1596
Db   331 GlyGlyLeuAlaLeuAlaLeuAlaArgGlyArgGlyArgGlyTrpSerValGlyValGlyAla 350

Qy  1597 --------CTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGTGCCTGCGCGGGTGATT 1644
                    :::|||             :::||||||
Db   351 ProGlyGlyValGlnIleGluThrTyrThrValProAlaAspLysCysGlyGlyLeuValIle 370

Qy  1645 GGCAAAGGTGGCAAGACCGTGAACGAACCTGCAGAACTTAACCAGTGCAGAAGTCATCGTG 1704
             |||||||:::          :::|||:::||
Db   371 GlyLysGlyGlyLysThrValAsnIleLysSerIleAsnGlnInSerGlyAlaHissValGluLeu 390

Qy  1705 CCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTTC 1764
             :::||||            :::
Db   391 GlnAsnArgAsnProProProAsnSerAspProAsnLeu--- 402

Qy  1765 TTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAGAATTGTACAACAGGTGAAGCAGCAG 1824
             :::||||||       :::||
Db   403 ------ArgArgPheThrIleArgGlyValProGlnInIleGluValAlaAla 417

Qy  1825 GAGCAGAAATACCC-----------TCAGGAGTCGCCTCACAGCGCAG 1862
             :::          ::::::||||||
Db   418 ArgGlnIleuIleAspGluLysValGlyLyThrAsnLeuGlyAlaProGlyAlaPheGly 437

Qy  1863 CAAGTGAGGCTCCCACGACCAGCAAAACGATGAGTAGCCCTTCCACACCT 1922
             |||           :::||
Db   438 GlnSerProPheSerGlnProProAlaProProHisGlnAsnThrPheProProArgSer 457

Qy  1923 GACAGAATGAGACCACAAACGCAGCCAGCCAGATCGGGAGCAACCA 1967
             :::::::||||||
Db   458 SerGlyCysPheProAsnMetAlaAlaLysValAsnPro 472
```

RESULT 3

```
ZBP2_CHICK
ID   ZBP2_CHICK              STANDARD;      PRT;     769 AA.
AC   Q8UVD9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zipcode-binding protein 2.
GN   ZBP2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 104-119; 242-267; 679-685; 705-711 AND
RP   743-766, FUNCTION, DEVELOPMENTAL STAGE, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21640447; PubMed=11781334;
RA   Gu W., Pan F., Zhang H., Bassell G.J., Singer R.H.;
RT   "A predominantly nuclear protein affecting cytoplasmic localization of
RT   beta-actin mRNA in fibroblasts and neurons.";
RL   J. Cell Biol. 156:41-52(2002).
CC   -!- FUNCTION: Binds to a 54-nucleotide localization signal (the
CC       zipcode) found in the 3' untranslated region of beta-actin mRNA
CC       and may play a role in its localization.
CC   -!- SUBCELLULAR LOCATION: Predominantly nuclear. Also detected in the
CC       cytoplasm.
CC   -!- DEVELOPMENTAL STAGE: The highest expression is found in 6-d
CC       embryos, is reduced to 30% before hatching and remains stable
CC       thereafter.
CC   -!- SIMILARITY: Contains 4 KH domains.
```

```
DR   EMBL; AF461020; AAL66365.1; -.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003730; F:mRNA 3' UTR binding; IDA.
DR   GO; GO:0008298; P:mRNA localization, intracellular; NAS.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat.
FT   DOMAIN     208    272       KH 1.
FT   DOMAIN     296    362       KH 2.
FT   DOMAIN     385    449       KH 3.
FT   DOMAIN     486    553       KH 4.
SQ   SEQUENCE   769 AA;  80644 MW;  875A6C83529969EB CRC64;

Alignment Scores:
Pred. No.:              5.89e-09        Length:         769
Score:                  260.50          Matches:        142
Percent Similarity:     36.0%           Conservative:   67
Best Local Similarity:  24.48%          Mismatches:     173
Query Match:            4.26%           Indels:         198
                                        Gaps:           31
                                        Db:

US-09-270-437D-6 (1-3412) x ZBP2_CHICK (1-769)

Qy   12 AGGCGAGGAGCGCCGGGTACCGGCCGGGGAGCCCGGGGCTCTCGGGGAAAGAGACGGAT   71
Db
Db  160 ArgSerGlyProValGlyAspProGlyProProGlyProProArgAla              173

Qy   72 GATGAACAAGCTTTACATGGGAACCTGAGCCCCGCCGACCTCCGGCA              131
Db
Db  174 GluArgGlyArgProProPro                                         181

Qy  132 GCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGC  191
Db
Db  182 AlaLeu         ThrGlyGlyAlaLeuProSer                          190

Qy  192 CTTCGTGGACTACCCGACCAGACTGGGC          CATCCG                  227
Db
Db  191 AlaAlaLeuPro  ProGlnLeuGlyProMetHisProProArgSerThrThr         208

Qy  228 CGCCATCGAGACCCTCTGCGGGTAAAGTGCATGGGAAAAATCATGGA              278
Db
Db  209 ValThrGluGluTyrArgGlyMetGlyMetValGlyLeuLeuIleIleGlyArgGlyGly  228

Qy  279 AGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCC  338
Db
Db  229 GluGlnIle      AsnLysIleGlnAspSerGlyCysLysValGln              243

Qy  339 TCCTCACCTGCAGTGGAGTGTTGGATGGAGTTTGGCTCAATATGGACAGTGGAGAA      398
Db
Db  244 IleSerProAspSerGlyGlyLeu                                      251

Qy  399 TGTGGAACAAGCAAAAATAGCCATGGAGAAGGAAGCTAAGCGGGCCATCAGTTGCAACGTGTCAACGTCACATATGCAACAAG  458
Db
Db  252 Pro                                                          252

Qy  459 AGAGAAGAAGCAAAATAGCCATGGAGAAGGAAGCTAAGCGGGCCATCAGTTGAGGAACTACTCCTT  518
Db
Db  253 GluArgGlySerValSerLeuThrGlySerProGluAlaValGlnLnLysAlaLysSerLeuMetLeu  272

Qy  519 CAAGATTTCCTACATCCCGATGAGGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCA  578
Db
Db  273 AspAspIleValSerArgArgGly    ArgGly                            281
```

```
Qy  579 GCGTGGGGACCACTCTTCCCGGGAGCAAGGCAC              GCCCCTGGGGCAC  626
Db  282                    GlyPro      ProGly  GlnPheHisAspTyrArgAlaAsnGlyGlnAsnGlyTh  297

Qy  627 TTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCTGGTCCCCACCCAGTTTGTTGG   686
Db  297 rValGln                       GluIleMetIleProLaGlyLysAlaGl     309

Qy  687 TGGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGCAGACCCAGTCCCG  746
Db  309 yLeuValIleGlyLysGlyGlyGluThrIleLysGlnLeuGlnLeuArgAlaGlyAlaGlyValLy  329

Qy  747 G  GTAGATATCCATAGAAAGAGAACTCTGGAGCTGCAGAGAGCCGTCACCATCCA        803
Db  329 sMetIlePheIleGlnAspGlySerGlnAsnThrAsnValAspLysProLeuArgIleI1    349

Qy  804 TGCCACCCCAGAGAGGGGACTTCTGAAGCATGCCGCATGGATTCTTGAAATCATCCAGAAAGA  863
Db  349 eGlyAspProTyrLysValGlnGlnAlaCysGluMetValMetAspIleLeu  ArgGl      368

Qy  864 GGCAGATGAG                       ACCAAACTAGCCGAAGAGAT          893
Db  368 uArgAspGlnGlyPheGlyAspArgAsnGlnuTyrGlySerArgIleGlyGlyGlyIl       388

Qy  894 TCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAG   953
Db  388 eAspValProValProArgHisSer       ValGlyValValIleGlyArgSerGlyGl   406

Qy  954 AAATTTGAAGAAAATTGAACATGAAACAGGAGACAAGGAGCCAAGATAACAATCTCATCTTTGCAGA  1013
Db  406 uMetIleLysIleLysGlnAsnAspAlaGlyValArgIleGlnPheLys       GlnAs       424

Qy 1014 TTTGAGCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGC  1073
Db  424 pAspGlyThr  GlyProGlnGluLysIleAlaHisIleMetGlyProGluArgCysGl     443

Qy 1074 CAGTGCTGAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATGATGCTGGC  1133
Db  443 uHisAlaAlaAlaArgIleIle                  AsnAspLeuLeuGl          454

Qy 1134 TGTTAACCAACAGCCAATCTGATCCCAGGGTTGAACCTGAGCGCACTTGGCATCTTTTC  1193
Db  454 n                                          GlnAs                454

Qy 1194 AACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCC          CGCGGAGCTCCCCCGC 1247
Db  455                SerLeuArgSerGlyProProGlyProProGlyHisGlyMetProProGl  471

Qy 1248 TGCCCCCTACCACCCCTTCACTACCACTCCGATACTTCTCCAGCCTGTACCCCATCA    1307
Db  471 yGly                          ArgGlyArgGlyArgGlyInGlyIleTrp    482

Qy 1308 CCAGTTTGGCCCGTTCCCGCCATCATCACTCCTATCCAGAGCAGGAGATTGTGAATCTCTT  1367
Db  483                           ProGlyGlyGlu  MetThrPheSe            492

Qy 1368 CATCCCAACCCAGGCTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCT   1427
Db  492 rIleProThrHisLysCysGlyLeuValIleGlyLysGlyLeuAsnValLysAlaIl     512

Qy 1428 GGGGAGATTCGGCCGAGCCTCTATCAAGATTGCC          CCTCCGGAAGGCCCAGA  1478
Db  512 eAsnGlnGlnArgGlyAlaPheValGluIleSerArgGlnLeuProProAsnGly  As     531

Qy 1479 CGTCAGGCGAAAGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGG  1538
Db  531 pProAsnPheLysLeuPheIleIleArgGlySerProGln                      544

Qy 1539 ACGGATCTTTGGGAAACTGAAAGGAGAAACTTCTTTAACCCCAAGAAGAAGAGTGAAGCT  1598
Db  545                       GlnIleGluHisAlaLysGlnProIleGluGlu  LysIl  557

Qy 1599 GGAAGCGCATATCAGAGTGCGCCTCTTCCACAGCTGGCCGGGTTGATTGGCAAAGGTGGC  1656
```

Db   557 eGluGly----------ProLeuCysProValGlyProGlyProGlyProGlyGly 572

RESULT 4
FUB2_HUMAN
ID   FUB2_HUMAN   STANDARD;   PRT;   707 AA.
AC   Q92945; O00301; Q9UNT5; Q9UQH5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 2 (FUSE binding protein 2) (KH
DE   type splicing regulatory protein) (KSRP) (p75).
GN   FUBP2 OR KHSRP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9605;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 72-85; 123-128; 267-281; 283-291;
RP   348-359; 472-486; 488-492; 620-644 AND 646-651, AND FUNCTION.
RC   TISSUE=Neuroblastoma, and Retinoblastoma;
RX   MEDLINE=97282621; PubMed=9136930;
RA   Min H., Turck C.W., Nikolic J.M., Black D.L.;
RT   "A new regulatory protein, KSRP, mediates exon inclusion through an
RT   intronic splicing enhancer.";
RL   Genes Dev. 11:1023-1036(1997).
RN   [2]
RP   SEQUENCE OF 1-115 AND 570-707 FROM N.A.
RX   MEDLINE=99189245; PubMed=8940189;
RA   Ring H.Z., Vameghi-Meyers V., Nikolic J.M., Min H., Black D.L.,
RA   Francke U.;
RT   "Mapping of the KHSRP gene to a region of conserved synteny on human
RT   chromosome 19p13.3 and mouse chromosome 17.";
RL   Genomics 56:350-352(1999).
RN   [3]
RP   SEQUENCE OF 60-707 FROM N.A., AND FUNCTION.
RC   TISSUE=B-cell lymphoma, and Skeletal muscle;
RX   MEDLINE=97094955; PubMed=8940189;
RA   Davis-Smyth T., Duncan R.C., Zheng T., Michelotti G., Levens D.;
RT   "The far upstream element-binding proteins comprise an ancient family
RT   of single-strand DNA-binding transactivators.";
RL   J. Biol. Chem. 271:31679-31687(1996).
RN   [4]
RP   FUNCTION, AND INTERACTION WITH PTBP1; PTBP2 AND HNRPH1.
RX   MEDLINE=20459250; PubMed=11003644;
RA   Markovtsov V., Nikolic J.M., Goldman J.A., Turck C.W., Chou M.-Y.,
RA   Black D.L.;
RT   "Cooperative assembly of an hnRNP complex induced by a
RT   tissue-specific homolog of polypyrimidine tract binding protein.";
RL   Mol. Cell. Biol. 20:7463-7479(2000).
CC   -!- FUNCTION: Binds to the dendritic targeting element and may play a
CC   role in mRNA trafficking (By similarity). Part of a ternary
CC   complex that binds to the downstream control sequence (DCS) of the
CC   pre-mRNA. Mediates exon inclusion in transcripts that are subject
CC   to tissue-specific alternative splicing. May interact with single-
CC   stranded DNA from the far-upstream element (FUSE). May activate
CC   gene expression.
CC   -!- SUBUNIT: Part of a ternary complex containing FUBP2, PTBP1, PTBP2
CC   and HNRPH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear. A small proportion is also found in
CC   the cytoplasm of neuronal cell bodies and dendrites (By
CC   similarity).
CC   -!- TISSUE SPECIFICITY: Detected in neural and non-neural cell lines.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to numerous
CC   frameshifts.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/

DR   EMBL; U94832; AAB53222.1; -.
DR   EMBL; AF093747; AAD29861.1; -.
DR   EMBL; AF093745; AAD29861.1; JOINED.
DR   EMBL; AF093748; AAD29862.1; -.
DR   EMBL; U69126; AAC50892.1; ALT_FRAME.
DR   Genew; HGNC:6316; KHSRP.
DR   MIM; 603445; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transport; mRNA transport; mRNA processing; mRNA splicing;
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; RNA-binding; Repeat.
FT   DOMAIN   144   208   KH 1.
FT   DOMAIN   233   299   KH 2.
FT   DOMAIN   322   386   KH 3.
FT   DOMAIN   422   489   KH 4.
FT   DOMAIN   569   682   4 X 12 AA IMPERFECT REPEATS.
FT   REPEAT   569   580   1.
FT   REPEAT   615   626   2.
FT   REPEAT   641   652   3.
FT   REPEAT   671   682   4.
FT   DOMAIN   7   67   GLY/PRO-RICH.
FT   DOMAIN   68   495   GLY-RICH.
FT   DOMAIN   496   610   ALA/GLY/PRO-RICH.
FT   CONFLICT   96   96   G -> V (IN REF. 2).
FT   CONFLICT   692   692   V -> G (IN REF. 2).
FT   CONFLICT   694   694   G -> GG (IN REF. 2).
FT   CONFLICT   707   707   Q -> A (IN REF. 2).
SQ   SEQUENCE   707 AA; 72709 MW; E07588DE43BCA8B6 CRC64;

Alignment Scores:
Pred. No.:   2.32e-08   Length: 707
Score: 251.00   Matches: 131
Percent Similarity: 35.95%   Conservative: 75
Best Local Similarity: 22.86%   Mismatches: 181
Query Match: 4.10%   Indels: 186
Db: 1   Gaps: 27

US-09-270-437D-6 (1-3412) x FUB2_HUMAN (1-707)

Qy   35 GCCGGGGAGCCCGGCCGCCGGCTCTCGGGGAGGAGACGGATGATGAACAAGCTTTACATCGGGA 94
              |||||||| |||||| ||| |||
Db   17 AlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGly----------AlaGly 27

Qy   95 ACCTGAGCCCCGCCGTCACCGCCGACC---- 124
              ||||||||||||||||||||||||
Db   28 GlyGlyProProGlyProProGlyAlaGlyGlyAspArgGlyGlyGlyProCysGly 47

Qy   125 ----------TCCGGCAGCTCTTTGGGACCACAGGAGGCGTCCCCTGGCGGGGGACCAGGTCC 172
                     |||||||::|| ||||||
Db   48 GlyGlyProGlyGlyGlySerAla-GlyGlyProSerGlnProProGlyGlyGlyPr 67

Qy   173 TGCTGAAGTCCGGCTACGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCCGCGCCA 232
              |||||| |||| ||| :::
Db   67 oGlyIleArgLys-AspAlaPheAlaAspAlaValGlnArgAlaArgGlnIleAlaAla- 86

Qy   233 TCGAGACCCTCCGGGTAAAGTGGAATTGCATGGG---- 267
              ::: |||||
Db   87 ----IleGlyIleGlyAspAlaAlaThrThrGlyAsnAsnSerThrProAspPheGlyP 105

Qy   268 ----------AAAATCATGGAAGTTGATTACTACCAGTCTCTAAAAGCTAAGGA 310
                       :::::::
Db   105 heGlyGlyGlnLysArgGlnLeuGluAspGlnProGluSerLysLysLeuAlaS 125

Qy   311 GCAGGAAATTCAGATTCGA----------AACATCCCTCCTCACCTG---- 348
              ||::| |||
Db   125 erGlnGlyAspSerIleSerSerGlnLeuGlyProIleHisProProArgProThrSerM 145

```
Qy    349  --------CAGTGGGGAGGTGTTGATGGGACTTTTGGCTCAATATGGGACAGTG----------  393
                    ::::|||||   ||||||||||::::
Db    145  etThrGluGlyTyrArgValProAspGlyMetVal--------GlyLeuIleIleGlyA  162

Qy    394  ---------GAGAATGTGGAACAAGTCAAACACAGACACAGAAACCGCCGTTGTCAACGTCA  445
                    |||:: ::::|||   |||::::
Db    162  rgGlyGlyGluGlnIleAsnLysIleGlnGlnAspSer--------------------  174

Qy    446  CATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTG  505
                    :: |||
Db    175  ----------GlyCysLysValGlnIleSerProAspSerGlyGly--------  186

Qy    506  AGAACTACTCCTTCAGATTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTCGCCCC  565
                    :::|||:::
Db    187  --------------LeuProGluArgSerValSerLeuThrGlyAlaP  198

Qy    566  CTCAGCGAGCCCAGCGTGGG--------------GACCACTCTTCCCGGGAGCAAGGCC  610
                    |||:::   |||   ::::||
Db    198  roGluSerValGlnLysAlaLysMetMetLeuAspAspIleValSerArgGlyArgGlyG  218

Qy    611  ACGCCCCTGG-----------------GGCACTTCTCAGGCCA----  637
                    |||||
Db    218  lyProProGlyGlnPheHisAspAsnAlaAsnGlyGlnThrValGln---  236

Qy    638  GACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCG  697
                    |||||:::|||
Db    237  --------------GluIleMetIleProAlaGlyLysLysLeuValIleG  250

Qy    698  GAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCC  757
                    |||||   |||   :::
Db    250  lyLysLysGlyGluThrIle--------------LysGlnLeuGluArgAlaGlyVaiL  266

Qy    758  ATAGA--------------AAAGAGAACTCTGGAGCTGCAGAGAGCCTGTCACCA  799
                    ::::|||   |||||:::
Db    266  ysMetIleLeuIleGlnAspGlySerGlnAsnThrAsn---ValAspLysProLeuArgI  285

Qy    800  TCCATGCCACCCCAGAGGGGACTTCTGAGCATGCGCATGATTCTTGAAATCATGCAGA  859
                    ||| :::::::::::|||:::
Db    285  ------ValGlyValValGlnGlnAlaCysGluMetValMetAspIleLeuArgA  305

Qy    860  AA-----------GAGGCAGATGAG--------ACCAAACTAGCCGAAGAGA  892
                    :: ||||||   |||
Db    305  snValThrLysAlaGlyPheGlyAspArgAsnGluTyrGlySerArgIleGlyGlyI  325

Qy    893  TTCCTTGAAAATCTGGCACACAATGGCCTTGGTTGGAAGACTGATTGGAAAAGAAGGCA  952
                    :::|||||| |||::: ||||||:::
Db    325  leAspValProValArgHisSer------ValGlyValValIleGlyArgSerGlyY  343

Qy    953  GAAATTTGAAGAAAAATTGAACATGAAACAGGGACCAAGATAACAATCCTCATCTTTCGCAGG  1012
                    :::||||||||   |:::::::|||   :::|||
Db    343  luMetIleLysLysIleGlnAsnAspAlaGlyValArgIleGlnPheLys-------GlnA  361

Qy    1013 ATTTGAGCATATACAACCCGGAAAGAAAACCATCACTGTGAAGGGGCACAGTTGAGGCCTGTG  1072
                    ::: |||   |||::: |||
Db    361  spAspGlyThr--GlyProGluLysIleAlaHisIleMetGlyProProAspArgCysG  380

Qy    1073 CCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGG  1132
                    ||||||:::::|||
Db    380  luHisAlaAlaArgIleIle--------------AsnAspLeuLeuG  391

Qy    1133 CTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTT  1192
                    |||
Db    391  ln------------------------  391

Qy    1193 CAACAGGACTGTCCGTGCATCTCCACCACGCAGGCCCGCGCGGAGCTCCCCCCGCTGCCC  1252
                    ::::   ||| |||||:::
Db    392  --------SerLeuArgSerGlyTyrProProGlyProGlyIleProLeuProG  408

Qy    1253 CCTACCACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGT  1312
                    : |||
Db    408  lyGlyArgGlyArgGlyArgGlyGly--------------AsnT  419

Qy    1313 TTGGCCCGTTCCCGACCATCATCACTCTTATCCAGAGCAGGAGAGATTGTGAATCTCTTCATCC  1372
```

```
Db    419  ::::|||||
           rpGlyProGlyGlyGlu--------------------MetThrPheSerIleP  430

Qy    1373 CAACCCAGGCTGCTGGGCGCCATCATCGGGAAGAAGAAGGGGCCACATCAAACAGCTGGCGA  1432
                    ||||||||||||||||||||||
Db    430  roThrHisGlyLysGlyLeuValIleGlyLysArgGlyGlyLysAsnValLysAlaIleAsnG  450

Qy    1433 GATTCGCCGGAGCCTCATCAAGATTGCC-----------CCTGCGGAGGCCCCAGACGTCA  1483
                    :: |||||||   :::::|||   |||||
Db    450  lnGlnThrGlyAlaPheValGluIleSerArgGlnLeuProGlnProThrGlyProThrS  470

Qy    1484 GCGAAAGATGGTCATCATCACCGGGCCACCGAA  1518
                    :::|||
Db    470  er---LysPheIleIleArgGlySerProGln  480
```

RESULT 5

```
PCB3_HUMAN
ID   PCB3_HUMAN      STANDARD;     PRT;     339 AA.
AC   P57721;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCBP3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67.301-316(2000).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF176329; AAG09240.1; -.
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8651; PCBP3.
DR   GO; GO:0005634; C:nucleus; ISS.
DR   GO; GO:0003677; F:DNA binding; ISS.
DR   GO; GO:0003723; F:RNA binding; NAS.
DR   GO; GO:0016071; P:mRNA metabolism; NAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       13     63       KH 1.
FT   DOMAIN       97    150       KH 2.
FT   DOMAIN      261    313       KH 3.
SQ   SEQUENCE   339 AA;  35938 MW;  F9F19FF7590C4188 CRC64;
```

Alignment Scores:

```
Pred. No.:                  4.27e-08       Length:              339
Score:                      245.50         Matches:              90
Percent Similarity:         38.65%         Conservative:         53
Best Local Similarity:      24.32%         Mismatches:          116
Query Match:                4.01%          Indels:              111
```

```
DB:

                    ||||||||     ::  ||||||||||  |||
Db      305  ------SerGluArgGlnIleThrGlyThrProAlaAsnIleSerLeuAlaGln 322

Qy     1537  GGACGGATCTTTGGGAAACTGAAAGAGGAA 1566
                  ||||||  ::   :::|||
Db      323  TyrLeuIleAsnAlaArgLeuThrSerGlu 332

RESULT 6
PCB3_MOUSE
ID   PCB3_MOUSE      STANDARD;      PRT;     339 AA.
AC   P57722;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCBP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics  and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by non-profit institutions as long  as its content is in no  way
CC   modified and this statement is not removed.  Usage by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AF176327; AAG09238.1; -.
DR   HSSP; Q07244; 1KHM.
DR   MGD; MGI:1890470; Pcbp3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       13     63       KH 1.
FT   DOMAIN       97    150       KH 2.
FT   DOMAIN      261    313       KH 3.
SQ   SEQUENCE   339 AA;  35958 MW;  A6F1C7C176A64F9C CRC64;

Alignment Scores:
Pred. No.:                              Length:           339
Score:            244.50                Matches:           90
Percent Similarity:  38.65%             Conservative:      53
Best Local Similarity: 24.32%           Mismatches:       116
Query Match:           3.99%            Indels:           111
DB:                    1                Gaps:              14

US-09-270-437D-6 (1-3412) x PCB3_MOUSE (1-339)

Qy    643  ATTGAATTCCCGCTGCGGATCCGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAG 702
                :::   ||::|||:||:::   :::   |||||:||||||||:||
Db     12  ValThrLeuThrIleArgLeuMetHisGlyLeuGluIleIleGlyLys 31

Qy    703  GAGGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCGGTAGATATCCATAGA 762
                :::||| |||  :::::::|| ::::   :|||||::||
Db     32  LysGlyGluThrValLysMetArgGluGluSerGlyAlaArgIleAsnIle------ 49

Qy    763  AAAGAGAACTCTGGAGCTGCGAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGACT 822
                |||::::||::||||||||
Db     50  --SerGlyAsnCysProGluArgIleValThrIleThrGlyProThrAspAlaIle 68

Qy    823  TCTGAAGCATGCCGCCATGATTCTT-----------GAAATCATGCAGAAAGAGGCA 867
                :::||||    ||||||             :::||:::   :::
Db     69  PheLysAlaPheAlaMetIleAlaTyrLysPheGluGluAspIleIleAsnSerMetSer 88

Qy    868  GATGAGACCAAACTAGCCGAA--GAGATTCCTCTGAAAATCTTGGCACACAATGGCTTG 924
                ||||
Db     89  AsnSerProAlaThrSerLysProProValThrLeuArgLeuValValProAlaSerGln 108

Qy    925  GTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTGAAGAAAATTGACATGAAACAGGG 984
                |||||||||||    :::||:::|||             |||||:|||
Db    109  CysGlySerLeuIleGlyLysGlyLySerLeuIleLeuGluLySGluLeuArgGluSerThrGly 128

Qy    985  ACCAAGATAACAATCTCATCTTTGCAGATTGAGCATATACAACCCGGAAAGAACCATC 1044
                :::||||||
Db    129  AlaGlnValGlnValAlaGly------AspMetLeuProAsnSerThrGluArgAlaVal 146

Qy   1045  ACTGTGAAGGCACAGTGAGGCCTGTGCCAGTGCTGAGATAGAGATTAATGAAGAAGCTG 1104
                ||||||: ||||||      :::|||
Db    147  ThrIleSerGlyThrProAspAlaIleIleGlnIleLeuGlnIleIleCysValValMet 166

Qy   1105  CGTGAGGGCCTTGAAAATGATATGCTGGCTGTGTTAACCACAGCCAATCTGATCCCAGGG 1164
                |||
Db    167  LeuGlu---------------------------------------------------- 168

Qy   1165  TTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCA 1224
                ||||||||
Db    169  ------------------------------------------SerProProLys 172

Qy   1225  GGGCCCGCGCGAGCTCCCCCGCTGCCCCTACCACCCC-------- 1263
                ||| ||                   |||
Db    173  Gly-------------AlaThrIleProTyrArgProLysProAlaSerThrProVal 187

Qy   1264  -----------------TTCACTACCCACTCCGATACTTCTCCAGCCTGTACCCC 1302
                   :::|||                            :::|||
Db    188  IlePheAlaAlaGlyGlyGlnAlaThrThrIleGlnIleGlyGlnTyr-------AlaIlePro 204

Qy   1303  CAT-------------------CACCAGTTTGGC-------CCGTTCCCGCAT 1329
                |||                   ||:|||
Db    205  HisProAspGlnLeuThrLysGlnLeuHisGlnLeuAlaMetGlnGlnThrProPheProPro 224

Qy   1330  ----------------------------------CATCACTCTTATCCA 1344
                ||||||
Db    225  LeuGlyGlnThrAsnProAlaPheProGlyGluLysProLeuHisSerSerGluGlu 244

Qy   1345  GAGCAGGAGATTGTG-------------------------GAAGGCCCA 1359
                ||||||| :::::
Db    245  AlaGlnAsnLeuMetGlyGlnSerGlyLeuAspAlaSerProAlaSerThrHis 264

Qy   1360  AATCCTCCATCCCAACCCAGCGTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATC 1419
                ||||||||
Db    265  GluLeuThrIleProAsnAspLeuIleGlyCysIleIleGlyLyArgGlnGlyThrLysIle 284

Qy   1420  AAACAGCTGCGAGATTGCGCAGCCTCTATCAAGATTGCCCTGG--GAAGGCCCA 1476
                ||||||| :::   |||   ||
Db    285  AsnGluIleArgGlnIleMetSerGlyAlaGlnIleLysIleIleAlaAsnAlaThrGluGlySer 304

Qy   1477  GACGTCAGCGAAAGGATGGTCATCACGGGTCATCACGGGTCATCACGGGCCACCCAGTTCAGGCCCAG 1536
                :::||| |||
Db    305  LysGlyGluThrValLysMetArgGluGluSerGlyAlaArgIleAsnIle------ 49
```

```
Qy   763 AAAGAGAACTCTGGAGCTGCAGAGAGCCTGTCACCATCCATGCCACCCAGAGGGACT 822
             ||||:  :::  ||  ||||||||
Db    50 ---SerGluGlyAsnCysProGluArgIleValThrIleThrGlyProThrAspAlaIle 68

Qy   823 TCTCAAGCACATGCCGCATGAGATTCTT------------GAAATCATGCAGAAAGAGGCA 867
             :::|| ||||||               ::||| ||::||
Db    69 PheLysAlaPheAlaMetIleAlaTyrLysPheGluGluIleIleAsnSerMetSer 88

Qy   868 GATGAGACCAAACTAGCCGAA---GAGATTCCTCTGAAAATCTTGGCACACAATGGCTTG 924
             :::    :::::       ::: ||| |||::::::::
Db    89 AsnSerProAlaThrSerLysProProValThrLeuArgLeuValValProAlaSerGln 108

Qy   925 GTTGAAGACTGATTGGAAAAGAAGGCAGAAAATTGAACAAAATTGAACATGAAACACGG 984
             ||  ||||||||||||||           :::||::||
Db   109 CysGlySerLeuIleGlyLysGlyGlySerLysIleLysGluIleArgGluSerThrGly 128

Qy   985 ACCAAGATACAAATCATCTTTGCAGGATTGAGCATATACACCCGAAAGAACCATC 1044
             ::: :::::     ||||||     ------    :::
Db   129 AlaGlnValGlnValAlaGly------AspMetLeuProAsnSerThrGluArgAlaVal 146

Qy  1045 ACTGTGAAGGGCACAGTTGAGGGCCTGTCCAGTGCTGAGATAGAGATTATGAAGAGCTG 1104
             :::::||  :::||  :::::|||     :::|||||||::::
Db   147 ThrIleSerGlyThrProAspAlaIlePheGlnCysValLysGlnIleCysValValMet 166

Qy  1105 CGTGAGGGCCTTTGAAAAATGATATGCTGGCGTGTTAACCAACAAGCCAATCTGATCCCAGGG 1164
             ||
Db   167 LeuGlu------------------------------------------------------ 168

Qy  1165 TTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCA 1224
             ||||||||
Db   169 ---------------------------------------SerProProLys 172

Qy  1225 GGGGCCCCGCGGGAGTCCCCCGCTGCCCCCTACCACCCC--------------------- 1263
             |||
Db   173 Gly----------------------AlaThrIleProTyrArgProLysProAlaCysThrProVal 187

Qy  1264 --------------------TTCACTACCACTCCGGATACTTCTCCAGCCTGTACCCC 1302
             ::::::     ||||
Db   188 IlePheAlaGlyGlyGlnAlaTyrThrIleGlnAlaGlnGlnTyr------------AlaIlePro 204

Qy  1303 CAT------------------------------------CACCAGTTTGGC--------CCGTTCCGCAT 1329
             |||                                    ||||||           ||||||
Db   205 HisProAspGlnLeuThrLysGlnLeuHisGlnLeuAlaMetGlnGlnThrProPhePreProPro 224

Qy  1330 -----------------------------------CATCACTCTTTATCCA 1344
                                                ||||||||
Db   225 LeuGlyGlnThrAsnProAlaPheGlyGluLysProGlyLysProLeuHisSerSerGluGlu 244

Qy  1345 GAGCAGGAGATTGTG--------------------------------------- 1359
             ||  :::::
Db   245 AlaGlnAsnLeuMetGlyGlnSerSerGlyLeuAspAlaSerProProAlaSerThrHis 264

Qy  1360 AATCTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGAAGAAGGGGCCACACATC 1419
             ||    ::||     |||||::::    :::|||||||::||
Db   265 GluLeuThrIleProAsnAspLeuIleGlyCysIleIleGlyArgGlnGlyThrLysIle 284

Qy  1420 AAACAGCTGCGCAGATTCGGCGAGCCTCTATCAAGATTGCCCCTGCG---GAAGGCCCA 1476
             :::::   ::::::||||||||:    :::||||     :::::||   ||
Db   285 AsnGluIleArgGlnMetSerGlyAlaGlnIleLysIleAlaAsnAlaThrGluGlySer 304

Qy  1477 GACGTCAGCGGAAAAGGATGGTCATCATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAG 1536
             ||||||       :::  :::     ||||||     |||::::||
Db   305 ------SerGluArgGlnIleIleThrIleThrGlyThrProAlaAsnIleSerLeuAlaGln 322

Qy  1537 GGACGGATCTTTGGGAAACTGAAGAGAGAA 1566
             ||||||||||
Db   323 TyrLeuIleAsnAlaArgLeuThrSerGlu 332
```

RESULT 7

FUB1_MOUSE

ID   FUB1_MOUSE   STANDARD;   PRT;   651 AA.
AC   Q91WJ8; Q8C0Y8;

DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 1 (FUSE binding protein 1) (FBP).
GN   FUBP1 OR D3ERTD330E.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Retina;
RX   MEDLINE=22382257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodriques S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE OF 1-286 FROM N.A. (ISOFORM 2).
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Regulates MYC expression by binding to a single-stranded
CC   far-upstream element (FUSE) upstream of the MYC promoter. May act
CC   both as activator and repressor of transcription (By similarity).
CC   -!- SUBUNIT: Interacts with SIAHBP1/FIR and JTV1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;

```
CC     Name=1;
CC        IsoId=Q91WJ8-1; Sequence=Displayed;
CC     Name=2;
CC        IsoId=Q91WJ8-2; Sequence=VSP_008322;
CC     Note=No experimental confirmation available;
-!- PTM: Ubiquitinated. This targets the protein for proteasome-
       mediated degradation (By similarity).
-!- SIMILARITY: Contains 4 KH domains.
CC
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     the European Bioinformatics Institute. There are no restrictions on its
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
CC
DR     EMBL; BC014763; AAH14763.1; -.
DR     EMBL; AK029458; BAC26457.1; -.
DR     MGD; MGI:1196294; D3Ertd330e.
DR     MGD; MGI:1924642; 9530027K12Rik.
DR     InterPro; IPR008160; Collagen.
DR     InterPro; IPR004087; KH_dom.
DR     InterPro; IPR004088; KH_type_1.
DR     Pfam; PF00013; KH; 4.
DR     SMART; SM00322; KH; 4.
DR     PROSITE; PS50084; KH_TYPE_1; 4.
KW     Transcription regulation; Trans-acting factor; Nuclear protein;
KW     DNA-binding; Repeat; Ubl conjugation; Alternative splicing.
FT     DOMAIN      95    160      KH 1.
FT     DOMAIN     181    247      KH 2.
FT     DOMAIN     271    335      KH 3.
FT     DOMAIN     372    439      KH 4.
FT     DOMAIN      16     21      POLY-GLY.
FT     DOMAIN     345    392      GLY-RICH.
FT     DOMAIN     446    556      PRO-RICH.
FT     VARSPLIC    67     67      D -> DGSWTNPSSTTHWEGMPSPFKD (in isoform
FT                                2).
FT                                /FTId=VSP_008322.
FT     CONFLICT    93     93      MISSING (IN REF. 2).
SQ     SEQUENCE   651 AA;  68539 MW;  F4EECA62FD9FA0D5 CRC64;

Alignment Scores:
Pred. No.:                       6.32e-08
Score:                           244.00
Percent Similarity:              32.57%
Best Local Similarity:           23.04%
Query Match:                     3.9%
DB:                              1
```

```
Length:      651
Matches:     162
Conservative: 67
Mismatches:  221
Indels:      255
Gaps:        34
```

```
US-09-270-437D-6 (1-3412) x FUB1_MOUSE (1-651)

Qy    404  AACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGA----   459
            ||    ||   ||||    ||      ||   |||||||:::
Db     78  AsnAspSerPheGlyAlaGlnLeuProMet-------HisGlnGlnSerArg          94

Qy    460  -----------GAAGAAGCAAAAATA----------                          474
                       ||||||    ||:::
Db     95  SerValMetThrGluGluTyrLysValProAspGlyMetValGlyPheIleIleGlyArg  114

Qy    475  GCCATGGAGAAGCTAAGCGGGCATCAGTTGAAACTACTCCTTCAAGATTTCCTACATC    534
            ||:::||:||||||    ||    |||:::
Db    115  GlyGlyGluGlnIleSerArgIleGlnGlnGluSer--GlyCysLysIleGlnIleAla  133

Qy    535  CCGGATGAAGAG-------------------GTGAGCTCCCCTTCGCCCCCT          567
            |||:::|||
Db    134  ProAspSerGlyLeuProGluArgCysMetLeuThrGlyThrProGluSerVal       153

Qy    568  CAGCGAGCCAGCTGGGGACCACTCTTCCCGGGAGCAAGGCCAC--GCCCCTGGG---    621
            ||:::||  |||||||||    |||:::
Db    154  GlnSerAlaLysArgLeuLeuAspGlnLeuValGlyLysGlyProAlaProGlyPhe   173

Qy    622  ------------GGCACTTCTCAGGCACGAGCAGATTGATTTCCCGCTGCCGGATCCTGGTC  669
                        |||||:::
```

```
Db    174  HisHisGlyAspGlyProGlyAsnAlaValGln---------GluIleMetIle    188

Qy    670  CCCACCCAGTTGTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACT    729
            |||:::||||||||||
Db    189  ProAlaSerLysAlaGlyLeuValIleGlyLysGlyGlyGluThrIle----------   204

Qy    730  AAGCAGACCCAGTCCGGGTAGATATCCATAGA---------AAAGAGAAC         771
            ||||||||:::
Db    205  LysGlnLeuGlnLeuArgAlaGlyAlaValLysMetValMetIleGlnAspGlyProGlnAsn  224

Qy    772  TCTGGAGCTGCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGGGACTTCTGAAGCA    831
            :::||:::||:::||||||||:::
Db    225  ThrGly---AlaAspLysProLeuArgIleThrGlyAspProTyrLysValGlnGlnAla  243

Qy    832  TGCCGCATGATTCTTGAAATCATGCGAAAGAGGCAGATAGACCAAACTAGCCGAAGAG    891
            |||||||:::||||||
Db    244  LysGluMetValLeuGluLeuIleLeuArgAspGlnGlyGlyGlyPheArgGluValArgAsnGlu  263

Qy    892  ATTCCTCTGAAAATCTTGCACACAATGGCTTG--------------------GTT    927
            ||:::||||||
Db    264  TyrGlySerArgIleGlyGlyGlyIleLeuAspValProIleProArgPheAlaVal   283

Qy    928  GGAAGACTGATTGGAAAAGAAGGCAGAAATTGAACATGAAGAAGCAGGGACC    987
            |||:::||||||:::||||||
Db    284  GlyIleValIleIleGlyArgAsnGlyGluMetIleLysLysIleGlnAspAlaGlyVal  303

Qy    988  AAGATAACAATCCATCTTTGCCAGGATTTGAGCATATACAACCCGGAAAGAACCATCACT   1047
            ||:::||||||
Db    304  ArgIleGlnPheLysProAspAspGlyThrThr----------ProAspArgIleAlaGln  320

Qy   1048  GTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGT   1107
            ||:::||:::||||||
Db    321  IleThrGlyProProAspArgCysGlnHisAla----AlaGluIleIle       335

Qy   1108  GAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAGCCAATCGATCCCAGGGTTG    1167
            |||:::||||||:::||||||
Db    336  ------------ThrAspLeuLeuArgSerValGlnAlaGlyAsn-----ProGly---  348

Qy   1168  AACCTGCAGCCGCACTTGCATCTTTCAACAGGACTTGCCGTGCATCTCCACCAGCAGGG   1227
            ||:::||||||
Db    349  ------------------------------------GlyProGlyProGly         353

Qy   1228  CCCCCGGGA--------------------GCTCCCCCGGCTGCCCCTCTAC   1257
            |||||||||
Db    354  GlyArgGlyArgGlyGlyGlyGlnGlyAsnTrpAsnMetGlyProProGlyGly----   371

Qy   1258  CACCCCTTCACTACCCCACCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGGTTGGC   1317
            |||||||||
Db    371                                                            371

Qy   1318  CCGTTCCGCATCATCACTCTTATCCAGAGCAGGAGGATTGTGAATCTCTTCATCCCAACC   1377
            ||:::||||||
Db    372  ------------------------------LeuGlnGluPhe---AsnPheValProThr   381

Qy   1378  CAGGCTGTGGGGCGCCATCATCGGGAAGAAGGGGGCCACACATCAACAGTTGGCAGATTC   1437
            |||:::||||||
Db    382  GlyLysThrGlyLysLeuIleIleGlyValGlyLysGlyGluThrIleLeuLysSerIleSerGlnGln   401

Qy   1438  GCCAGGGCCCTTCATCAAGATTGCCCCTGCGGAAGGCCCACAGGTCAGCGAAGATGGTC   1497
            ||:::||:||||||:::||||||
Db    402  SerGlyAlaArgIleGluLeuGlnLeuArgSerProProAsnAlaAspProAsnMetLys  421

Qy   1498  ATCATCACC----------------------------------------------------   1506
            ||:::|||
Db    422  LeuPheThrIleArgGlyThrProGlnGlnIleAspTyrAlaArgGlnLeuLeuGluGlu  441

Qy   1507  ----GGGCACC---------GGAAGCCCAGTT--                          1526
                |||||||
Db    442  LysIleGlyGlyProValAsnProLeuysGlyProProHisGlyVal   461

Qy   1527  ---------------CAAGGCCCCAGGGACGGACGATTCTTTGGAAACTGAA         1559
                           ||||||||:::
```

Db 462 ProGlyProHisGlyProProGlyProGlyProGlyThrProMetGlyProTyrAsn 481

Qy 1560 AGAGGAAAACTTCTTTAACCCCAAAGAAGAAGAAGTGAAGCTGGAAGCGCA--------- 1607
                      ||::|
Db 482 ProAla---------ProTyrAsnProGlyProGlyProAlaProHisGlyPro 497

Qy 1608 --------TATCAGAGTGCCCTCTTCCACAGCTGCCCGGTGATTGGCAAAGTGGCAA 1658
                    ::: ||| |||||
Db 498 ProAlaProTyrAlaPro--------GlnGlyTrpGlyAsnAlaTyrProHisTrpGln 514

Qy 1659 GACCGTGAACGAACTGCAGAACTGCAGAGTCATCGTGCCCTGTGACCAAAC 1718
                  ::: |||
Db 515 GlnGlnAla--------------------------------ProPro 519

Qy 1719 GCCAGATGAAAATGAGGAAGTGATCGTCAGAATATCGGGCACTTCTTTGCTAGCCAGAC 1778
                 |||:::::|| |:::
Db 520 AspProAlaLysAlaGlyAlaAlaAsp--------------------ProAsn 529

Qy 1779 TGCCACAGCG-----------------CAAGATCAGGGAAATTGTACAACA 1811
                 ||| |||
Db 530 SerAlaAlaTrpAlaAlaAlaTyrTyrAlaHisTyrTyrGlnGlnAlaGlnProProPro 549

Qy 1812 GGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCGCCAGCAAGTGAGG 1871
                 ||||||| ||| :::
Db 550 AlaAlaProAlaGlyAla-----ProAlaThrThrGlnThrAsnGlyGlnGlyAspGln 567

Qy 1872 CTCCCACAGCCACCAGCAAAACAA-------------------- 1895
                 |||||||
Db 568 GlnAlaProAlaProAlaGlyGlnValAspTyrThrLysAlaTrpGluGluTyrTyrLys 587

Qy 1896 CGGATGAATGTAGCCCTTCCAACACCTGACAGAGAATGAGACCAGCCAGCCAGATC 1955
                 :::|||| ||:::|||| ||| :::
Db 588 LysMetGlyGlnAlaValProAlaPro----------AlaGlyAlaProPro 601

Qy 1956 GGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTG 2015
                 |||||| |||:::
Db 602 GlyGlyGlyGlnProAspTyrSerAlaAlaTrpAlaAlaTrpAlaGluTyrTyrArgGlnGlnAlaAlaTyr 621

Qy 2016 CCGGAGCCCCTGAGAACCCCAGGGCGGGGCCGAGGAGGGGGCGGGAAGTCAGGTTGCCA 2075
                                                 ||||||
Db 622 TyrAlaGlnThrSerProGln--------------------GlyMetPro 631

Qy 2076 GAACCACCGAGCCCCCGCCTCCGGCCCCCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTC 2135
                 :::||| |||||||||||||||| |||
Db 632 Gln-His--------ProProAlaProGlnGlyPhe-AlaAsnHisAlaArgSerHisH 648

Qy 2136 ACCAT 2140
                 |||||
Db 648 isHis 649

RESULT 8
FUB1_HUMAN
ID   FUB1_HUMAN   STANDARD;   PRT;   643 AA.
AC   Q96AE4; Q12828;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Far upstream element binding protein 1 (FUSE binding protein 1) (FBP)
DE   (FBP1).
GN   (DNA helicase V) (HDH V).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 191-193; 203-206;
RP   272-279; 284-290; 300-314; 321-328; 394-397; 409-411; 430-438 AND
RP   440-443, AND FUNCTION.
RC   TISSUE=Leukemia;
RX   MEDLINE=94170991; PubMed=8125259;
RA   Duncan R., Bazar L., Michelotti G., Tomonaga T., Krutzsch H.,
RA   Avigan M., Levens D.;
RT   "A sequence-specific, single-strand binding protein activates the far
RT   upstream element of c-myc and defines a new DNA-binding motif.";
RL   Genes Dev. 8:465-480(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 45-64; 133-145; 271-283; 308-321; 379-386; 414-424 AND
RP   430-439.
RX   MEDLINE=21127960; PubMed=11222755;
RA   Vidigni A., Ochem A., Triolo G., Falaschi A.;
RT   "Identification of human DNA helicase V with the far upstream
RT   element-binding protein.";
RL   Nucleic Acids Res. 29:1061-1067(2001).
RN   [4]
RP   PARTIAL SEQUENCE, AND MASS SPECTROSCOPY.
RX   MEDLINE=22166132; PubMed=12176931;
RA   Rappsilber J., Ryder U., Lamond A.I., Mann M.;
RT   "Large-scale proteomic analysis of the human spliceosome.";
RL   Genome Res. 12:1231-1245(2002).
RN   [5]
RP   INTERACTION WITH SIAHBP1.
RX   MEDLINE=20337922; PubMed=10882074;
RA   Liu J., He L., Collins I., Ge H., Libutti D., Li J., Egly J.-M.,
RA   Levens D.;
RT   "The FBP interacting repressor targets TFIIH to inhibit activated
RT   transcription.";
RL   Mol. Cell 5:331-341(2000).
RN   [6]
RP   INTERACTION WITH JTV1, UBIQUITINATION, AND PROTEASOME-MEDIATED
RP   DEGRADATION.
RX   MEDLINE=22716800; PubMed=12819782;
RA   Kim M.J., Park B.-J., Kang Y.-S., Kim H.J., Park J.-H., Kang J.W.,
RA   Lee S.W., Han J.M., Lee H.-W., Kim S.;
RT   "Downregulation of FUSE-binding protein and c-myc by tRNA synthetase
RT   cofactor p38 is required for lung cell differentiation.";
RL   Nat. Genet. 34:330-336(2003).
RN   [7]
RP   STRUCTURE BY NMR OF 278-447 IN COMPLEX WITH SINGLE STRANDED FUSE DNA.
RX   MEDLINE=21864720; PubMed=11875576;
RA   Braddock D.T., Louis J.M., Baber J.L., Levens D., Clore G.M.;
RT   "Structure and dynamics of KH domains from FBP bound to
RT   single-stranded DNA.";
RL   Nature 415:1051-1056(2002).
CC   -!- FUNCTION: Regulates MYC expression by binding to a single-stranded
CC   far-upstream element (FUSE) upstream of the MYC promoter. May act
CC   both as activator and repressor of transcription.
CC   -!- SUBUNIT: Interacts with SIAHBP1/FIR and JTV1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q96AE4-1; Sequence=Displayed;

```
CC      Name=2;
CC        IsoId=Q96AE4-2; Sequence=VSP_008321;
CC        Note=No experimental confirmation available;
CC   -!- PTM: Ubiquitinated. This targets the protein for proteasome-
CC       mediated degradation.
CC   -!- SIMILARITY: Contains 4 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U05040; AAA17976.2; -.
DR   EMBL; BC017247; AAH17247.1; -.
DR   PDB; 1J4W; 06-MAR-02.
DR   Genew; HGNC:4004; FUBP1.
DR   MIM; 603444; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
KW   Transcription regulation; Trans-acting factor; Nuclear protein;
KW   DNA-binding; Repeat; Ubl conjugation; Alternative splicing;
KW   3D-structure.
FT   DOMAIN       99    163       KH 1.
FT   DOMAIN      184    250       KH 2.
FT   DOMAIN      274    338       KH 3.
FT   DOMAIN      375    442       KH 4.
FT   DOMAIN       13     26       GLY-RICH.
FT   DOMAIN      348    395       GLY-RICH.
FT   DOMAIN      449    559       PRO-RICH.
FT   VARSPLIC    642    643       GQ -> CRFDPASIELAL (in isoform 2).
FT                               /FTId=VSP_008321.
FT   CONFLICT     96     96       Q -> QS (IN REF. 1).
SQ   SEQUENCE   643 AA;  67473 MW;  086D4EAA0ACF807B CRC64;
```

```
Alignment Scores:
Pred. No.:            7.84e-08              Length:        643
Score:               242.50                Matches:       145
Percent Similarity:   35.48%               Conservative:   81
Best Local Similarity: 22.76%              Mismatches:    241
Query Match:           3.96%               Indels:        170
DB:                                        Gaps:           27

US-09-270-437D-6 (1-3412) x FUB1_HUMAN (1-643)

Qy  336 CCCTCCTCACCTGCCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGGACAGTGGA 395
            ||||||     :::      |||||||
Db   10 ProSerSerGlySerAlaGlyThrGlyGlyGly---------------GlyGlyGly     22

Qy  396 GAATGTGGAACAAGTCAACACAGACAAACCGCCGTTGTCAACGTCACATATGCAAC    455
            |||||    ||||    :::    |||::::
Db   23 GlyGlyGly-GlyValAsnAspAlaPheLysAspAlaLeuGlnArg------AlaAr     39

Qy  456 AAGAGAAGAGCAAAAATAGCCATGGAGAGCTAAGCCGGGCATCAGTTTGAGAACTACTC   515
            :::    ||||||||    :::     :::
Db   39 gGlnIleAlaAlaLysIleGlyGlyAspAlaGlyThrSerLeuAsnSerAsnAspTyrGl   59

Qy  516 CTTCAAGATTTCTACATCCCGATGAAGGAGGTGAGTCCCCTTCGCCCCTTCAGCGAGC    575
            :::       |||  |||    :::
Db   59 yTyrGlyLysArgProLeuGluAspGlyAspGlnProAspAlaLysLysValAl        79

Qy  576 CCAGCGTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCGGGGGCACTTCTCAGGC    635
            ::  |||          ::
Db   79 aProGlnAsnAspSerPheThrGln------LeuProProMetHisGlnGlnGl         96

Qy  636 CAGACAGATTGATTTCCCGCTGCCGGATCCTGCGCCTGTCCCCACCCAGTTTGTTGGTGCCATCAT 695
            ||                  :::
Db          ::
```

```
Db   96 nArgSerVal---MetThrGluTyrLysValProAspGlyMetValGlyPheIleIl    115
Qy  696 CGGAAGGAGGGCTTGACCATAAGAACATCACTAGCAGACCCAGTCCCGGTAGATAT     755
           |||||::::   |           ::::::
Db  115 eGlyArgGlyGlyGluGlnIleSerArgIleGlnGlnGlySerGlyLysIleGlnIl    135

Qy  756 CCATAGAAAAGAGAACTCTGGAGCTGCGAGAGAAGCCTGTCACCATCGCACCCCAGA   815
           |             :::      |||||
Db  135 e---AlaProAspSerGlyGlyLeuProGluArgSerCysMetLeuThrGlyThrProGl 154

Qy  816 GGGGACTTCTGAGCATGCCGCATGATTCTTGAAATCATGCAGAAA-----------    861
           |||       |||    ::::::   :::||||:::::|||
Db  154 uSerValGlnSerAlaLysArgLeuLeuAspGlnIleValGluGlyProAlaPr     174

Qy  862 ---------GAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGC   911
Db  174 oGlyPheHisHisGlyAspGlyProGlyAsnAlaValGln--------GluIleMetIl   191

Qy  912 ACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGAGGCAGAAAATTGAAGAAAATTGA 971
           :::|||||||   |||||||
Db  191 eProAlaSerLysAlaGlyLeuValIleGlyLysGlyGlyGluThrIleLysGlnLeuGl   211

Qy  972 ACATGAAACACGGGACCAAGATAACATCTCATCTTTGCAGGATTTGAGCATATACAACCC  1031
           :::||||||   |||||
Db  211 nGluArgAlaGlyValLysMetValMet------IleGlnAsp---------GlyPr    225

Qy 1032 GGAAAAGAACCATCACTGTGAGGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGAT 1091
           |::::   |||||||
Db  225 oGlnAsnThr--------------GlyAlaAsp---------                  231

Qy 1092 TATGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCCAA 1151
           |||    ||||
Db  232 ---LysProLeuArg------IleThrGlyAspProTyrLysValGlnGlnAlaLysGl  248

Qy 1152 TCTGATCCCAGGGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCT 1211
           :::::   :::|||||
Db  248 uMetVal------LeuGluLeuIl                                    254

Qy 1212 ATCTCCACCAGCAGGGCCCCGGGAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTAC    1271
           |||
Db  254 eArgAspGlnGlyGlyPheArgGluVal---------                       263

Qy 1272 CCACTCCGGATACTTCTCCAGCTGTACCCCGTTCCCGTTCCCGCATCA            1331
           ::::::::||   |||
Db  264 --------ArgAsnGluTyrGly--------                             268

Qy 1332 TCACTCTTATCCAGACAGCAGAGATTGTGAATCTTCTTCATCCCAACCCCAGCGTGTGGGCGC 1391
           |||||||   |||
Db  269 -SerArgIleGlyGlyAsnGluGlyIleAspValProIleProArgPheAlaValGlyIl  288

Qy 1392 CATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGAGTTCGCCGGAGCCTCTAT 1451
           ::::|||   :::::::
Db  288 eValIleGlyArgAsnGlyGluMetIleLysLysIleGlnAsnAspAlaGlyValArgIl  308

Qy 1452 CAAGGATTGCCCCTGGGAAGGCCCCAGAGCTCAGCGAAGGATAGTGTCATCATCACCGGGCC 1511
           ::::::::::|||
Db  308 ---ArgAsnGluTyrGly---ThrThrProGluArgIleAlaGlnIleThrGlyPr     327

Qy 1512 ACCGGAA---GCCCAGTTCAAGGCCCAG                                   1536
           |||   |||::
Db  327 oProAspArgCysGlnHisAlaAlaGluIleIleThrAspLeuLeuArgSerValGlnAl  347

Qy 1537 ---------GGACGGATCTTTGGGAAA                                    1554
Db  347 aGlyAsnProGlyLysProGlyLeuGlnGluProGlyArgGlyArgGlyLysAsnTr     367

Qy 1555 --------------CTGAAAGAGGAAAAACTTCTTAACCCCAAAGAAGAAGT          1592
Db  367 pAsnMetGlyProGlyLeuGlnGluProAsnPheIle--------               381

Qy 1593 GAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGCCCGGTGATTGGCAAAGG   1652
           ::::::   :::|||||
Db  382 --------ValProThrGlyLysThrGlyLeuIleGluIleGlyLysLysGl        394
```

```
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding; Repeat.
FT   DOMAIN        13     75       KH 1.
FT   DOMAIN        97    162       KH 2.
FT   DOMAIN       279    343       KH 3.
SQ   SEQUENCE    356 AA;  37497 MW;  6D1A26127GCA206D CRC64;
```

```
Alignment Scores:
Pred. No.:                13                    Length:         356
Score:                    238.50                Matches:         89
Percent Similarity:       41.78%                Conservative:    66
Best Local Similarity:    23.99%                Mismatches:     137
Query Match:               3.90%                Indels:          79
DB:                                             Gaps:            13
```

```
US-09-270-437D-6 (1-3412) x PCB1_RABIT (1-356)

Qy   619 GGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGTCGCGAGATCCTGGTCGCGGATCCAGG 678
             ||| |||:||||:|:|:||         :::||||||||          ::||:
Db     4 GlyValThrGluSerGlyLeuAsnValThrLeuThrIleArgLeuLeuMetHisGlyLys  23

Qy   679 TTTGTTGGTGCCATCATCGGAAAGACGAGGGCTTGACCATAAAGAACATCACTAAGCAGACC 738
             |||||:|||:|||||||||:|:|||||    :|||:||| ::|:|||
Db    24 GluValGlySerIleIleGlyLysLysGlySerValLysSerArgIleArgGluGluSer   43

Qy   739 CAGTCCCGGTGAGTATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGCCTGTCACC 798
             :::         ::|         ||:
Db    44 GlyAlaArgIleAsnIle-------SerGluGlyAsnCysProGluArgIleIleThr   60

Qy   799 ATCCATGCGCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATCATGCAG 858
             :::||             ::|||          :::::
Db    61 LeuThrGlyProThrAsnAlaIlePheLysAlaMetIleIleIleAspLysLysLeuGlu  80

Qy   859 AAAGAGGCAGATGAGACCAAACTA------------GCCGAAGAGATTCCTCTG 900
             ::::|:     :::                    |||
Db    81 GluAspIleAsnSerSerMetThrAlaAlaSerArgProProValThrLeu  100

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGAGGCAGAAATTTG 960
             ::::::::      |||      ::: ||| :::
Db   101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysLysGlyLysLysIle  120

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
             ||||| :::::|||        ::|||::
Db   121 LysGluLeuArgGluSerThrGlyAlaGlnValAlaGlnValAlaGly----AspMetLeu 138

Qy  1021 ATATACAACCCGGAAAGAACCATCACTGTGGAAGGGCACAGTTGAGGCTGTGCCAGTGCT 1080
             ::        :::    :::          |||||         ::::::
Db   139 ProAsnSerThrGluArgAlaIleThrIleGluAlaGlyValProGlnSerValThrGluCys 158

Qy  1081 GAGATAGAGATT---------------ATGAAGAAGCTGCGTGAGCCTTGAAAATGATATG 1128
             :::::|||             :::::   ||| :::
Db   159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal  178

Qy  1129 CTGGCTGTTAACCACACAGCCAATCTGATCCCAGGGTTGAACCTGCCACTTGGCATC 1188
             |||||          ::: ||||||
Db   179 MetThrIleProTyrGlnPro------------MetProAlaSerSerProValIle 193

Qy  1189 TTTTCAACAGGACTGTCCGTGCTATCTCCACAGCAGGCCCCGCGGAGCTCCCCCGCT 1248
             |||||          :::||||
Db   194 CysAlaGlyGlyGlnAspArgCysSerSerAspAlaAlaGly----------- 206

Qy  1249 GCCCCCTACCACCCCTTCACTACCAC------------TCCGGTACTTC 1287
             ::|||            ||||
Db   207 -----------TyrProHisAlaThrHisAspLeuGluGlyProLeuAspAlaTyr--- 222

Qy  1288 TCCAGCCTGTACCCCCCATCACCAGTTTGGCCCG------------ 1320
             ||||:           :::::
Db   223 --SerIleGlyGlnHisThrIleSerProLeuAspLeuAlaLysLeuAsnGlnVal 241

Qy  1321 ---------------TTCCCGCATCATCAC------- 1335
                            :::
Db   242 AlaArgGlnGlnSerHisPheAlaMetMetHisGlyGlyThrGlyPheAlaGlyIleIleAsp 261

Qy  1336 ---TCTTATCCAGAG------------CAGGAGATT 1356
```

```
Qy  1653 TGGGCAAGACCGTGACGAACTGCAGAACTTAACAGTGCAGAGAGTCATCGTGCCTCGTGA 1712
             |||||:||||::|:       :::       ::: |||:::  ||||:|
Db   394 yGlyGluThrIleLysSerIleSerGlnGlnSerGlyAlaArgIleGluArgGlnArgAs 414

Qy  1713 CCAAACGCCAGATGAAAATGAGGAAGTG---ATCGTCAGAATTATCGGGCAC--------- 1761
             :||::|            :|:       :::  |||||
Db   414 nProProAsnAlaAspProAsnMetLysPheThrIleArgGlyThrProGlnGl 434

Qy  1762 ---TTCTTGCTAGCCAGACTGCACACGCGCAAGATCAGGGAAATTGTACAACAGGTGAA 1817
             :::|||            :::|||||           |||
Db   434 nIleAspThrAlaArgGlnLeuLeuGluLeuGlyLysGlyGlyValProValAsnProLeuGl 454

Qy  1818 GCAGCAGGAGGAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCA 1877
             |||||
Db   454 yProProValProHisGlyProHisGlyVal-ProGlyProHisGlyProGlyProP 474

Qy  1878 CAGGCACCAGCAAAAACAACGATGAATGTAGCCCTTCCACACACCTGACAGAATGAGACCA 1937
             ||:|||                               |||   :::
Db   474 roGly-ProGly----------------ThrPro------MetGlyPro 482

Qy  1938 AACGCAGCCAGATCGGAGCAAAACAAAGACCATCTGAGGAATGAGAAGTCTGCGG 1997
             :::                       :::
Db   483 TyrAsnProAlaProTyrAsnProGlyPro--------- 492

Qy  1998 AGGGGCGCCAGGGACTCTGCCAGG---CCCTGAGAACCCCAGGGGCCGAGGAGGGGCGGG 2054
             |||            |||         :::|||       |||
Db   493 ------ProGlyProAlaProHisGlyProProAlaProGlnGlyTyrAlaProGlnGlyTrpGly 510

Qy  2055 GAAGGTCAGCCAGGTTGCCAGAA------CCACCGAGCCCCGCC 2093
             :::       |||||         |||||
Db   511 AsnAlaTyrProHisTrpGlnGlnGlnAlaProProAspProAla 525
```

```
RESULT 9
PCB1_RABIT
ID   PCB1_RABIT   STANDARD;      PRT;   356 AA.
AC   O19048;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1).
GN   PCBP1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207122; PubMed=10101190;
RA   Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT   "Tissue-specific translational regulation of alternative rabbit
RT   15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL   Nucleic Acids Res. 27:1828-1836(1999).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ003023; CAA05814.1; -.
DR   HSSP; Q07244; 1KHM.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
```

```
Db    262  SerSerSerProGluValLysGlyTyrTrpAlaSerLeuAspAlaSerThrGlnThrThr  281
                                                              |||||
Qy   1357  GTGAATCCTTCATCCAACCAGCTGTGGCGCCCATCATCGGGAAGAAGGGGCCACAC  1416
            |||    |||||           :::||   :::::||||||||:||||||
Db    282  HisGluLeuThrIleProAsnAsnLeuIleGlyCysIleIleGlyIleGlyArgGlnGlyAlaAsn  301

Qy   1417  ATCAAACAGCTGGCGAGATTCGCCGGAGCCTCATCAAGATTGCC---CCTGCGGAAGGC  1473
            :  :::|||||       |:    || |||||  ::|
Db    302  IleAsnGluIleArgGlnMetSerGlyAlaGlnIleIleLysIleAlaAsnProValGluGly  321

Qy   1474  CCAGACGTCAGCGGAAGAGATGGTCATCATCACCCGGCCACCGGAAGCCCAGTTCAAGGCC  1533
            |||       |||||||||        |||||||||        :::
Db    322  Ser------SerGlyArgGlnValThrIleThrGlySerAlaAlaSerIleSerLeuAla  339

Qy   1534  CAGGGACGACGGATCTTTGGGAAAACTGAAAGAGGAA  1566
            |||      |||    :::||||      |||
Db    340  GlnTyrLeuIleAsnAlaArgLeuSerSerGlu  350
```

RESULT 10
Y475_ARATH

```
ID   Y475_ARATH     STANDARD;      PRT;    606 AA.
AC   P58223; O49507; Q8H0Y7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Putative nucleic acid binding protein At4g18375.
GN   At4g18375 OR F28J12.2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaugnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Marra M.A., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14591172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda A., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vayberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P58223-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P58223-2; Sequence=VSP_008899, VSP_008900;
CC         Note=May be due to a competing acceptor site. No experimental
CC         confirmation available;
CC   -!- SIMILARITY: Contains 5 KH domains.
CC   -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
CC       gene model prediction. At4g18370 and At4g18375 were originally
CC       fused into a single gene.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL021710; CAA16717.1; ALT_SEQ.
DR   EMBL; AL161548; CAB78839.1; ALT_SEQ.
DR   EMBL; AY133701; AAM91635.1; -.
DR   EMBL; BT001108; AAN64172.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 5.
DR   SMART; SM00322; KH; 5.
DR   PROSITE; PS50084; KH_TYPE_1; 5.
KW   Hypothetical protein; Nuclear protein; RNA-binding; Repeat;
KW   Alternative splicing.
FT   DOMAIN       35     99       KH 1.
FT   DOMAIN      138    210       KH 2.
FT   DOMAIN      311    380       KH 3.
FT   DOMAIN      394    455       KH 4.
FT   DOMAIN      535    599       KH 5.
FT   VARSPLIC    532    532       L -> F (in isoform 2).
FT                               /FTId=VSP_008899.
FT   VARSPLIC    533    606       Missing (in isoform 2).
FT                               /FTId=VSP_008900.
SQ   SEQUENCE   606 AA;    65760 MW;    61F135BBB8647C0C CRC64;
```

```
Alignment Scores:
Pred. No.:              1.39e-07        Length:         606
Score:                  238.50          Matches:        105
Percent Similarity:     38.40%          Conservative:   92
Best Local Similarity:  20.47%          Mismatches:     216
Query Match:            3.90%           Indels:         100
DB:                     1               Gaps:           16

US-09-270-437D-6 (1-3412) x Y475_ARATH (1-606)
```

```
Qy   577  CAGGCTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCC  636
Db   3    GluArgLysGlyLysArgLysGlnIleIleGlnArgAsnAsnSerGluSerAsnArgAsnGlnLys  22

Qy   637  AGACAGATTGATTCCCG----------CTGCGGATCCTGGTC  669
Db   23   ArgArgIleSerHisAspLysArgIleAsnArgAspGluLeuValValTyrArgIleLeuLeuCys  42

Qy   670  CCCACCCAGTTGTTGGTGCCATCATCGGGAAAGGAGGGGCTTGACCATAAAGACATCACT  729
Db   43   ProIleAspValValGlyLysValValIleGlyLysLysSerGlyLysValIleAsnAlaIleArg  62

Qy   730  AAGCAGACCCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGACCTGCAGAGAAG  789
Db   63   HisAsnThrLysAlaLysIleLysValPhe---AspGlnLeuHisGlyCysSerGlnArg  81

Qy   790  CCTGTCACCATCGCACCCAGAGGGACTTCTGAGCAGCCATGCCGCCATGATTCTTGAA  849
Db   82   ValIleThrIleTyrCysSerSerVallysGluLysGlnGlnGlu----------IleGly  97

Qy   850  ATCATCAGAAGAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTG  909
Db   98   PheThrLysSerGluAsnGluProLeuCysCysAlaGlnAsnAspAlaLeuLeuLysVal--  116

Qy   910  GCAACAAATGGCTTGGTT----------  927
Db   117  --TyrAspAlaIleValAlaLeuSerAspSerGluGluAsnAsnThrLysThrAsnValAspArg  135

Qy   928  GGAAGACTGATTGGA  942
Db   136  AspAspAsnLysGluCysArgLeuLeuValProPheSerGlnSerSerLeuGly  155

Qy   943  AAAGAAGGGCAGAAATTTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCA 1002
Db   156  LysAlaGlyGluAsnIleLysAsnIleLysArgIleLeuArgArgThrAlaSerValLysValVal  175

Qy   1003 TCTTTGCAGGATTTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAAGGGCACAGTT 1062
Db   176  Ser--LysAspValSer----AspProSerHisValCysAlaMetGluTyrAspAsn  192

Qy   1063 GAGGGCCCTGTGCCAGTGCCTGATAGAGATTATGAAGAAGCTGCCTGAGGCCCTTTGAAAAT 1122
Db   193  ValValValIleSerGlyGluProGluSerValLysGlnAlaLeuPheAlaValSerAla  212

Qy   1123 GATATGCTGGCTGTGTAACCAACAAGCCAATCTG----------ATC 1158
Db   213  IleMetTyrLysIleAsnProArgGluAsnIleProLeuAspSerThrSerGlnAspVal  232

Qy   1159 CCAGGGTTGAACCTC----------AGCGCACTTGGC 1185
Db   233  ProAlaAlaSerValIleValProSerAspLeuSerAsnSerValTyrProGlnThrGly  252

Qy   1186 ATCTTTTCAACAGGACTGTCCGTGTATCTCCACCAGCAGGGGCCGCCGGAGCTGCCCCCC 1245
Db   253  PheTyrSerAsnGlnAspHisIleLeuLeuGlnGlnGlyAlaGlyValProSerTyrPheAsn  272

Qy   1246 GCTGCCCCCTACCACCCCTTCACTACCCCACTCCGA----------  1281
Db   273  AlaLeuSerValSerAspPheGlnGlyTyrAlaGluThrAlaAlaAsnProValProVal  292

Qy   1282 TACTTCTCCAGCCTGTACCCCATCACCAGTTTCGCCCCGTTCCGCCATCACTCTTAT 1341
```

```
Db   293  PheAlaSerLeuProValThrHisGlyPheGly----------GlySerSerArg  308
Qy   1342 CCAGCAGGAGGATTGTGAATCTTCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGG 1401

Db   309  SerGluGluValPheLysValLeuCysProLeuCysAsnIleMetArgValIleGly  328
Qy   1402 AAGAGGGGGCACACATCAAACAGCTGGCAGATTCGCCGGAGCCTCTATCAAGATTGCC 1461

Db   329  LysGlyGlySerThrIleLysArgIleArgCysIleSerCysIleGluValAsn  348
Qy   1462 CCTGCGAA----------GGCCAGAGCGTCAGCGAAGGATGTCATCATCACCGGGCCA 1512

Db   349  AspSerArgThrLysCysGlyAspAsp------GluCysValIleLeuValThrAlaThr  366
Qy   1513 CCGGAAGCCAGTTCAAGGCCGGACGGATCTTTGGGAAACTGAAAGAGGAAAACTTC 1572

Db   367  GluSerProAspAspMetLysSerMetAlaValGluAlaValLeuLeuLeuGlnGluTyr  386
Qy   1573 TTTAACCCCAAAGAAGAAGAGTGAAGCTGGAAGGCCATATCAGATGCGCCTCCTTCCACAGCT 1632

Db   387  IleAlaAsnAspGluAspAlaGluAsnVallysMetGlnLeuLeuValSerGlyLysValIle  406
Qy   1633 GGGCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACCGCAGAACTTAACCAGTGCA 1692

Db   407  GlyCysValIleGlyLysSerGlySerValIleAsnGluIleArgLysAsnThrAsnAla  426
Qy   1693 GAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATT 1752

Db   427  AsnIleCysIleSerLysLysGlyGlnGlyGlyGlu----------LeuValGluVal  441
Qy   1753 ATCGGGCACTTCTTTGTCTAGCCAGACTGCACAGCGCAGATCAGGGAAATTGTACACAG 1812

Db   442  SerGlyGluVal--------------SerSerValArgAspAlaLeuLeuIleGln  454
Qy   1813 GTGAAGCAGCAGAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCACAGCCAAGTGAGGC 1872

Db   455  Ile-ValLeuValArgLeuArgGluArgGluLeuArgGlyAspLysAspValAlaThrArgLy 474
Qy   1873 TCCCACAGGCACCCAGCAAAACACGGATGAATGTAGC 1909

Db   474  sProProAla--------ArgThrAspAsnCysSer  483
```

```
RESULT 11
NOA2_HUMAN
ID   NOA2_HUMAN      STANDARD;       PRT;    492 AA.
AC   Q9UNW9; O43267; Q9UEA1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RNA-binding protein Nova-2 (Neuro-oncological ventral antigen 2)
DE   (Astrocytic NOVA1-like RNA-binding protein).
GN   NOVA2 OR ANOVA OR NOVA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20197319; PubMed=10735272;
RA   Ueki K., Ramaswamy S., Billings S.J., Mohrenweiser H.W., Louis D.N.;
RT   "ANOVA, a putative astrocytic RNA binding protein gene that maps to
RT   chromosome 19q13.3.";
RL   Neurogenetics 1:31-36(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99007301; PubMed=9789075;
RA   Yang Y.Y., Yin G.L., Darnell R.B.;
RT   "The neuronal RNA-binding protein Nova-2 is implicated as the
RT   autoantigen targeted in POMA patients with dementia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:13254-13259(1998).
RN   [3]
RP   SEQUENCE OF 29-492 FROM N.A.
```

```
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Brower A., Sanders C., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT   D19S412.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 405-480.
RX   MEDLINE=99148126; PubMed=10368286;
RA   Lewis H.A., Chen H., Edo C., Buckanovich R.J., Yang Y.Y.-L.,
RA   Musunuru K., Zhong R., Darnell R.B., Burley S.K.;
RT   "Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding
RT   domains.";
RL   Structure 7:191-203(1999).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 405-491 IN COMPLEX WITH RNA.
RX   MEDLINE=20139984; PubMed=10676814;
RA   Lewis H.A., Musunuru K., Jensen K.B., Edo C., Chen H., Darnell R.B.,
RA   Burley S.K.;
RT   "Sequence-specific RNA binding by a Nova KH domain: implications for
RT   paraneoplastic disease and the fragile X syndrome.";
RL   Cell 100:323-332(2000).
CC   -!- FUNCTION: May regulate RNA splicing or metabolism in a specific
CC       subset of developing neurons (By similarity). Binds single strand
CC       RNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: Brain. Expression restricted to astrocytes.
CC   -!- DOMAIN: The third KH domain (KH3) recognizes specifically 5'-
CC       UCAY-3'.
CC   -!- DISEASE: Autoantigen in the paraneoplastic opsoclonus myoclonus
CC       ataxia (POMA), a paraneoplastic neurological syndrome/disorder
CC       (PNS/D) associated with breast cancer, fallopian cancer, and SCLCa
CC       and characterized primarily by loss of inhibitory control of motor
CC       neurons in the spinal cord and brainstem. Recognized by the IgG
CC       autoantibody ANNA-2 (also called anti-Ri).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U70477; AAB86611; ALT_INIT.
DR   EMBL; AF083898; AAC72355.1; -.
DR   EMBL; AC006540; AAD13116.1; -.
DR   PDB; 1DTJ; 18-FEB-00.
DR   PDB; 1EC6; 07-APR-00.
DR   Genew; HGNC:7887; NOVA2.
DR   MIM; 601991; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Antigen; Nuclear protein; RNA-binding; Repeat; 3D-structure.
FT   DOMAIN       10     26       BIPARTITE NUCLEAR LOCALIZATION SIGNAL
FT                                (POTENTIAL).
FT   DOMAIN       32     99       KH 1.
FT   DOMAIN      130    196       KH 2.
FT   DOMAIN      232    405       ALA-RICH.
FT   DOMAIN      332    393       GLY-RICH.
FT   DOMAIN      406    473       KH 3.
FT   CONFLICT    247    247       A -> R (IN REF. 1).
FT   CONFLICT    265    267       PAA -> TAT (IN REF. 1).
SQ   SEQUENCE   492 AA;  49008 MW;  41B63EAF689256B CRC64;
```

```
Alignment Scores:
Pred. No.:                1.41e-07     Length:       492
Score:                    238.00       Matches:      124
Percent Similarity:       35.85%       Conservative: 61
Best Local Similarity:    24.03%       Mismatches:   216
Query Match:              3.89%        Indels:       116
DB:                                    Gaps:         16

US-09-270-437D-6  (1-3412)  x  NOA2_HUMAN  (1-492)

Qy    559  TCGCCCCCTCAGCGAGCCAGCGTGGGACCACCTCTTCCCGGAGCAAGGCCACGCGCCCT   618
                ::|||||||::                  :::::|||:::||||
Db     15  ThrProProGluValValCysThrLysArgSerAsnThrGlyGluGluGly--------   31

Qy    619  GGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCACCCAG   678
                                            :::::|||:::|||::::|||||
Db     32  --------------------------GluTyrPheLeuLysValLeuIleProSerTyr   42

Qy    679  TTTGTTGGTGCCATCATGCGGAAAAGGAGGGCTTGACCATAAAGAACCATCAAGCAGACC   738
                |||:|||||||||||||||          :::::|||:::::|||||||
Db     43  AlaAlaGlySerIleIleGlyLysGlyGlnThrIleValGlnLeuGlnLysGluThr    62

Qy    739  CAGTCCCGGTAGATATCCATAGAAAAGAGAAC----------TCTGGAGCTGCAGAAGCCT   792
                :::  :::  :::  :::  :::        ||  |:|
Db     63  GlyAlaThrIleLysLeuSerLysSerLysAspPheTyrProGlyThrThrGluArgVal   82

Qy    793  GTCACCATCCATGCCACCCCCAGGAGGGG-----------ACTTCTGAAGCA   831
                ::                              ::|||
Db     83  CysLeuValGlnGlyThrAlaGlyThrAlaGluAlaLeuAlaValHisSerPheIleAlaValGluGluLys   102

Qy    832  TGCGCCATGCATGATTCTTGAAATCATGCAGAAA----   861
                |||                              |||
Db    103  ValArgGluIleProGlnAlaMetThrLysProGluValValAsnIleLeuGlnProGln   122

Qy    862  --------GAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCA   912
                      |||                         |||::::::
Db    123  ThrThrMetAsnProAspArgAlaLysGlnAla------------------LysLeuLeuVal   137

Qy    913  CACAATGGCTTGGTTGGAAGACTGATTGGAAAAAGAGGCAGAAAATTGAAGAAAATTGAA   972
                |||                    :::|||||
Db    138  ProAsnSerThrAlaGlyLeuIleLleGlyLysGlyGlyAlaThrValLysGlyAlaValMet   157

Qy    973  CATGAAACAGGGACCAAGATAACAATCCATCTTTCCAGGATTTGAGCATATACAACCCG   1032
                :::::|||:||||               :::  :::|||
Db    158  GluGlnSerGlyAlaTrpValGlnAsnLeuSerGlnLysProGlu---GlyIleAsnLeuGln   176

Qy   1033  GAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCTGAGATAGAGATT   1092
                |||                         |||
Db    177  GluArgValValThrValSerGlyGluProGluGlnValHisLysLysAlaValSerAlaIle   196

Qy   1093  ATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGATCGGTGTGTTAACCACAAGCCAAT   1152
                :::::||||||:::||                   :::::
Db    197  ValGlnLysValGlnGluAspProGlnSerSerCysLeuAsnIleSerTyrAlaAsn   216

Qy   1153  CTGATC--CCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTG   1209
                :::               |||           :::
Db    217  ValAlaGlyProValAlaAsnSerAsnProThrGly------------------   228

Qy   1210  CTATCTCCACCAGCACAGGGCCCCGGGAGCTCCCCCGCTGGCCCCTTACCACCCCTTCACT   1269
                |||||                    :::||:::
Db    229  ---SerProTyrAlaSerProAlaAspValLeuProAlaAlaAlaAlaAlaSerAlaAlaAla   247

Qy   1270  ACCCACTCCGGATAC------------TTCTCAGCCTG   1296
                |||||               |||||
Db    248  AlaAlaSerGlyLeuLeuGlyProAlaGlyLeuAlaGlyValGlyAlaPhe-ProAlaAl   267

Qy   1297  TACCCCC------------   1303
                |||
Db    267  aLeuProAlaPheSerGlyThrAspLeuLeuAlaIleSerThrAlaLeuAsnThrLeuAl   287

Qy   1304  ------------ATCACCAGTTTGGCCGTTCCCG   1326
                                ::::  |||  |||
```

```
Db    287  aSerTyrGlyTyrAsnThrAsnSerLeuGlyLeuGlyLeuAsnSerAlaAlaAlaSerGl   307
Qy   1327  CATCATCACTCTTATCCAGAGCAGAGAGATTGTGAATCTTCATCCCAACCCAGGCTGTG  1386
                ::: :::       :::::::                          |||
Db    307  yValLeuAlaAlaValAlaAlaG:yAlaAsnProAlaAlaAlaAlaAlaAlaAsnLeuLe   327

Qy   1387  GGCGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGCGAGATTCGCCGGAGCC  1446
                :::||   ||
Db    327  uAlaSerTyrAlaGlyGluAlaGlyAlaGlyProAlaGlyGlyAlaAlaProProProPr   347

Qy   1447  T--------CTATCAAGATTGCCCCTGGGAGGCCCAGGAGGGCCAGACGTCAGCGAAAGG  1491
                       ||
Db    347  oProProGlyAlaLeuGlySerPheAlaLeuAlaAlaAlaAlaAsnGlyTyrLeuGl    367

Qy   1492  ATGGTCATCATCACC--------------GGGCCACCGGAAGCCCAGTTCAAGGC  1532
           |||||                          ||||||
Db    367  yAlaGlyAlaGlyGlyGlyAlaGlyGlyGlyProLeuValAlaAlaAlaAlaAl   387

Qy   1533  CCAGGACGGATCTTTGGGAAACTGACGAGGAAAACTTCTTTAACCCGAAAGAGAGT  1592
                :||| ||      |||
Db    387  aAlaGlyAlaAlaAlaGlyAlaGlyPheLeuThrAlaGlyLeuLysLeu------AlaAlaGluSerAl   405

Qy   1593  GAAGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAAGG  1652
           |||       :::             ||||||||
Db    405  aLysGluLeuValGluIleAlaValProGluAsnLeuvalGlyAlaIleLeuGlyLysGl   425

Qy   1653  TGGCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAGTC--ATCGTGCCTCG  1709
           |||||||::::      :::         ::
Db    425  yGlyLysThrLeuValGluTyrGlnGluLeuThrGlyAlaArgIleGlnIleSerLysLysVa  445

Qy   1710  TGACCAAACGCCAGATGAAAATGAGGAAGTGATCCTCGAGAATTATCGGGCACTTCTTTGC  1769
           ;    :::        |||         ||
Db    445  sGlyGluPheLeuProGlyThrArgAsnArgArgValThrIleThrGlySerProAlaAl   465

Qy   1770  TAGCCAGACTGCACACGACGCAAGATCAGGGAAATTGTACACACAGGTGAAGCAGCAGGAGCA  1829
           |:::||  ||||||       |||
Db    465  aThrGlnAlaAlaGlnTyrLeuIle------------------SerGlnArgVa   477

Qy   1830  GAAATACCCTCAGGGAGTC---GCCTCACAGCGCAGCAAGTGAGGC  1872
           |||  ||           ||||
Db    477  lThrGlyTyrGluGlnGlyValArgGlyAlaSerAsnProGlnLysValGly   492
```

```
RESULT 12
NOA1_RAT
ID   NOA1_RAT      STANDARD;      PRT;    474 AA.
AC   Q80WA4;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1)
DE   (Fragment).
GN   NOVA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BDIX; TISSUE=Cerebellum;
RA   Knudsen A., Monstad S.E., Vedeler C.A.;
RT   "Nova-1, the paraneoplastic Ri antigen, is associated with breast
RT   cancer.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   FUNCTION.
RX   MEDLINE=22120219; PubMed=12124753;
RA   Kumar D.V., Nighorn A., St John P.A.;
RT   "Role of Nova-1 in regulating alpha2N, a novel glycine receptor
RT   splice variant in developing spinal cord neurons.";
RL   J. Neurobiol. 52:156-165(2002).
CC   -!- FUNCTION: Functions to regulate alternative splicing in neurons by
CC       binding pre-mRNA in a sequence-specific manner to activate exon
CC       inclusion. It binds specifically to the sequence UCAUY (By
```

```
CC       similarity). Acts to regulate a novel glycine receptor alpha-2
CC       chain splice variant (alpha-2N) in developing spinal cord.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 3 KH domains.
```
```
DR   EMBL; AY262017; AAP20872.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
KW   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Antigen; Nuclear protein; RNA-binding; Repeat.
FT   DOMAIN         26     42       BIPARTITE NUCLEAR LOCALIZATION SIGNAL
FT                                  (POTENTIAL).
FT   DOMAIN         48    115       KH 1.
FT   DOMAIN        146    212       KH 2.
FT   DOMAIN        248    384       ALA-RICH.
FT   DOMAIN        396    463       KH 3.
FT   NON_TER       474    474
SQ   SEQUENCE      474 AA;  48443 MW;  70D0926B5BECD131 CRC64;
```

```
Alignment Scores:
Pred. No.:                                Length:            474
Score:                 2.02e-07           Matches:           103
Percent Similarity:     235.50            Conservative:       60
Best Local Similarity:  35.43%            Mismatches:        162
Query Match:            22.39%            Indels:            135
DB:                      3.85%            Gaps:               14
                            1
```

```
US-09-270-437D-6 (1-3412) x NOA1_RAT (1-474)

Qy    562  CCCCCTCAGGCGAGCCCAGCGTGGGACCACTCTTCCCGGAGCAAGGCCACGCCCCTGGG   621
               |||||||          ::::|||                          |||   ::
Db     21  ProProAspSerArgLysArgProLeuGluAlaProProGluAlaProGlyAlaGly---SerThrLys   39

Qy    622  GGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGGGGATCCTGGTCTCCCACCCAGTT   681
               ||::::             :::   |||:::||||   ||:::|||||
Db     40  ArgThrAsnThrGlyGluAspGlyAlaTyrPheLeuValLeuValIleProSerTyrAla   59

Qy    682  GTTGGTGCCATCATCGGAAGAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAG   741
               ||||||||||               |||      |||:::|||:::||
Db     60  AlaGlySerIleIleGlyLysGlyGlnThrIleValGlnLeuValGlnLysGluThrGly   79

Qy    742  TCCCGGGTAGATATCCATAGAAAAGAGAAC------TCTGGAGCTGCAGAGGAGCCTGTC   795
               ::: :::         ::: :::::||       |||:::
Db     80  AlaThrIleLysLeuSerLysLysSerLysAspPheTyrProGlyThrProGluArgValCys   99

Qy    796  ACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATCATG   855
               ||       ::           :::       ||:::    :::
Db    100  LeuLeuGlnIleThrIleGluAlaLeuAsnAlaValHisGlyPheIleAlaGluLysIle   119

Qy    856  CAGAAAGAGCCAGAGGAGACGACCAAACTAGCCGAGAGATTCCTCTG-----------   900
               :::::        ||||:::   ::::::|||:::
Db    120  ArgGluMetProGlnAsnValAlaLysThrGlnValSerIleLeuGlnProGlnThr   139

Qy    901  -------------------------------------AAAATCTTGGCACACAATGGCTTCGTTGGA   930
                                                    ||||||::: :::
Db    140  ThrValAsnProAspArgIleLysGlnValLysIleIleValProAsnSerThrAlaGly   159

Qy    931  AGACTGATTGGAAAAGAAGGAGGCAGAAAATTGAACAGAAAATTGAACAGGAACAGGAACAAAG   990
               :::       ||||||::::
Db    160  LeuIleIleGlyLysGlyGlyGlnThrIleValGlnAlaThrValAlaIleMetGluGlnSerGlyAlaTrp   179

Qy    991  ATAACAATCTCATCTTTGCAGGATTTGAGCATATACACCCGGAAAGAACCATCACTACTG  1050
               :::       :::|||                    |||   :::||||||
```

```
Db   180  ValGlnLeuSerGlnLysProAsp---GlyIleAsnLeuGlnGluArgValValThrVal  198
Qy  1051  AAGGGCACAGTTGAGGCCTGCCAGTGCCAGTAGAGATTATGAAGAAGCTGCGTGAG      1110

Db   199  SerGlyGluProGlnAsnArgLysAlaValGluLeuIleLeuGlnLysIleGlnGlu     218
Qy  1110  ...                                                           1110

Db   219  AspProGlnSerGlySerCysLeuAsnIleSerTyrAlaAsnValThrGlyProValAla  238
Qy  1110  ...                                                           1110

Db   239  AsnSerAsnProThrGlyPheProTyrAlaAlaAsnThrAlaGluValAlaLeuProThrAlaAla  258
Qy  1111  ...                  ---GCCTTGAA---                           1119

Db   259  AlaAlaAlaGlyLeuLeuLeuGlyHisAlaAsnLeuAlaGlyValAlaAlaAlaPheProAlaVal  278
Qy  1120  --AATGATATGCTCGCTGTTAACCAACAGCCAATCTGATCCCA                   1161

Db   279  LeuSerGlyPheThrGlyAsnAspLeuValAlaIleThrSerAlaLeuAsnThrLeuAla  298
Qy  1162  ---GGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCT     1215

Db   299  SerTyrGlyTyrAsnLeuAsnThrLeuGlyLeu-------GlyLeuSer---------    312
Qy  1216  CCACCAGCACAGGCCCCGCCGGAGCT---------------CCCCCGCT            1248

Db   313  --GlnAlaAlaAlaAlaThrGlyAlaLeuAlaAlaAlaAlaAlaAlaAlaSerAlaAsnProAlaAla  331
Qy  1249  GCCCCCTACCACCCCTTCACTACCCAC---------                          1275

Db   332  AlaAlaAlaAsnLeuLeuAlaThrTyrAlaAlaSerGluAlaSerGlySerThrAla    351
Qy  1276  ...                              ---TCC                       1278

Db   352  GlyGlyThrAlaGlyThrPheAlaLeuGlySerLeuAlaAlaAlaAlaAlaThrAlaAlaAlaThrAsn  371
Qy  1279  GGATACTTCTCCAGCGTGTACCCCCATCACCAGTTGGCCGTCCCGCATCACTCT      1338

Db   372  GlyTyrPheGlyAlaAlaSerProLeuAlaAlaSerAlaIleLeuGlyThrGluLysSer  391
Qy  1339  TATCCA---GAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGCGCCATC  1395

Db   392  ThrAspGlySerLysAspValValGluIleAlaValProGluAsnLeuValGlyAlaIle  411
Qy  1396  ATCGGGGAAGAAGGGGGCCACACATCAAACACAGCTGGCGAGATTCGGCGGAGCCTCTATCAAG  1455

Db   412  LeuGlyLysGlyGlyLysThrLeuValGluThrTyrGlnGluLeuThrArgIleGln    431
Qy  1456  ATTTGCCCCTGCGGAAGGC-------CCAGACCTCAGCGGAAGGATGGTCATCATCACC  1506

Db   432  IleSer--LysLysGlyPheValProGlyThrArgAsnArgLysValThrIleThr    450
Qy  1507  GGGGCACCGCAAGCCCAGTTCAAGGCGACGGATCTTTGGGAAACTGAAAGAGAGAA     1566

Db   451  GlyThrProAlaAlaThrGlnAlaGlnAsnTyrLeuIleThrGlnAsnArgIleThrTyrGlu  470
```

RESULT 13
PBP2_YEAST

ID   PBP2_YEAST   STANDARD;   PRT;   413 AA.
AC   P38151;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PAB1-binding protein 2.
GN   PBP2 OR YBR233W OR YBR1531.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Dubois E., el Bakkoury M., Glansdorff N., Messenguy F., Pierard A.,
RA   Scherens B., Vierendeels F.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PARTIAL CHARACTERIZATION.
RX   MEDLINE=99038243; PubMed=9819425;
RA   Mangus D.A., Amrani N., Jacobson A.;
RT   "Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-
RT   binding protein, regulates polyadenylation.";
RL   Mol. Cell Biol. 18:7383-7396(1998).
CC   -!- SUBUNIT: Interacts with PAB1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z36101; CAA85196.1; -.
DR   PIR; S46109; S46109.
DR   GermOnline; 138776; -.
DR   SGD; S0004437; PBP2.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT   DOMAIN   66   130   KH 1.
FT   DOMAIN   148   213   KH 2.
FT   DOMAIN   330   394   KH 3.
SQ   SEQUENCE   413 AA; 45782 MW; 638846509BCE1840 CRC64;

Alignment Scores:

Pred. No.:	2.09e-07	Length:	413
Score:	235.00	Matches:	102
Percent Similarity:	40.20%	Conservative:	56
Best Local Similarity:	25.95%	Mismatches:	151
Query Match:	3.84%	Indels:	84
Db:	1	Gaps:	16

US-09-270-437D-6 (1-3412) x PBP2_YEAST (1-413)

```
Qy   523  ATTTCCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTGGCCCCTCAGCGAGCCCAGCGT  582
Db    10  IleThrThrThrProThrThrValLeuValSerProAsnThrLeuLysProAsnThr...Arg--LysLys  28

Qy   583  GGGGGACCACCTCTTCCCGGGACGAA                                        606
Db    29  GlyGluAspThrSerGluGluGlnLeuGluAlaGluLeuAlaLeuAlaLeuLysLysAsp       48

Qy   607  ------GGCCACGCCCCTGGGGCCACTTCCAGGCCAGACAGATT------GATTCCCG         654
Db    49  AlaAspAspSerHisSerAsnAspHisAspSerProAspAsnValProSerAspValHis       68

Qy   655  CTGGCGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACC      714
Db    69  LeuThrMetLeuCysLeuValLysHisAlaSerLeuIleValGlyHisLysGlyAlaThr       88

Qy   715  ATAAAGAACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCT        774
Db    89  IleSerArgIleLysSerGluThrSerArgAlaGlyIleAsnIleSerAsnAsnIleArg--    107

Qy   775  GGAGCTGCAGAAGCCGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGC         834
Db   108  GlyValProGluArgIleValThrValArgGlyThrCysAspAspValAlaLysAlaTyr      127

Qy   835  CGCATGATTCTTGAAATCATGCAGAAAGAGACCAGAAGACCAAACTAGCCGAA---          888
```

Db 128 GlyMetIleValArgAlaLeuLeuGluGluHisGlyAsnGlyGluAspIle 147
        |||||||::: :::|||
Qy 889 GAGATTCCTTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAA 948

Db 148 GluIleSerIleAsnLeuLeuIleProHisHisLeuMetGlyCysIleIleGlyLysArg 167
        ||| ::: |||::||||
Qy 949 GGCAGAAATTGAAGAAAATTGAACATGAACAGGACCAAGATAACAATCTCATCTTTG 1008

Db 168 GlySerArgLeuArgGluIleGluAspAlaLeuAlaLysLeuPheAlaSerProAsn 187
        ||::|:::::|||||||
Qy 1009 CAGGATTTGAGCATATACAACCCCGAAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCC 1068

Db 188 Gln------LeuLeuLeuSerAsnAspArgIleLeuThrIleAsnGlyValProAspAla 205
        ::: ::: |||::|| |||:::||||
Qy 1069 TGTGCCAGTGCTGAGATAGACATTATGAAGAAGCTG------------------------ 1104

Db 206 IleHisIleAlaThrPheTyrIleSerGlnThrLeuLeuAsnPheGlnMetGluSerPro 225
        |||||||
Qy 1105 --------------------CGTGAGCCTTGAA------------AATGATATGCTGGCT 1134

Db 226 GlnLysAsnValLysArgSerIleTyrTyrGlnProThrGlnPheAsnSerValLeuIle 245
        ||| :::
Qy 1135 GTTACCAACAAGCCAATCTGATC------------------IleSerSerThrLeu 282

Db 246 AspHisSerGlnProAsnThrIlePheHisGlnArgAsnHisGlnTyrHisProSerAsp 265
        ||||| |||||
Qy 1159 ------CCAGGGTTGAACCTGACGGCCACTTGGCATCTTTCAACAGGA 1200

Db 266 LysLeuLeuSerTyrLysProAsnLysAsnLeuPro------IleSerSerThrLeu 282
        ||||| ||||||
Qy 1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGGAGCTCCCCCGCTGCCCCCTACCAC 1260

Db 283 LeuSerMetAlaThrProGlnTyrThrThrAlaSerValAlaAsnAlaThrAlaPheGln 302
        |||||||::: ::||||
Qy 1261 TTCACTACCCACCCGGATACTTCTCCAGCCTGTACCCC 1302

Db 303 ProAsnPheValIleProAsnValThrValLeuAspGlyProValIleSer--------- 319
        |||||||
Qy 1303 TCTTATCCAGAGCAGGAG 1353

Db 320 ------ProAlaProGlyAsnHisLeuLeuMetAsnPheValGlnGlnGlu 334
        ::||| |||||||
Qy 1354 ATTGTGAATCTTCATCCCAACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGAGGGGGCA 1413

Db 335 Ile------PheIleAspGluLysPheValGlyAsnValIleGlyLysGlyGlyLys 351
        ::: :::||||||||
Qy 1414 CACATCAAACAGCTGCGAGATTCGCCGGAGCCTCTATCAAGATTGCC---CCTGCGGAA 1470

Db 352 HisIleAsnSerValLysGluSerThrGlyCysSerIleIleIleGluGlnAspProValGlu 371
        :: |||||
Qy 1471 GGCCCAGACGTCAGCCAAAGGATGGTCATCATCACCGGG 1509

Db 372 GlySer------SerGluArgArgLeuThrIleArgGly 382
        ||| |||||| ||| |||
Qy 1471 ... SerGluArgArgLeuThrIleArgGly 382

RESULT 14
PCB1_HUMAN

ID   PCB1_HUMAN   STANDARD;   PRT;   356 AA.
AC   Q15365; Q13157; Q14975;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1) (Nucleic acid
DE   binding protein SUB2.3).
GN   PCBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9531278; PubMed=7607214;
RA   Leffers H., Dejgaard K., Celis J.E.;

RT   "Characterisation of two major cellular poly(rC)-binding human
RT   proteins, each containing three K-homologous (KH) domains.";
RL   Eur. J. Biochem. 230:447-453(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9016208; PubMed=7556077;
RA   Kiledjian M., Wang X., Liebhaber S.A.;
RT   "Identification of two KH domain proteins in the alpha-globin mRNP
RT   stability complex.";
RL   EMBO J. 14:4357-4364(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=9420381O; PubMed=8152927;
RA   Aasheim H.-C., Loukianova T., Deggerdal A., Smeland E.B.;
RT   "Tissue specific expression and cDNA structure of a human transcript
RT   encoding a nucleic acid binding [oligo(dC)] protein related to the
RT   pre-mRNA binding protein K.";
RL   Nucleic Acids Res. 22:959-964(1994).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC   preferentially to oligo dC.
CC   -!- SUBCELLULAR LOCATION: Loosely bound in the nucleus. May shuttle
CC   between the nucleus and the cytoplasm.
CC   -!- TISSUE SPECIFICITY: Abundantly expressed in skeletal muscle,
CC   thymus and peripheral blood leucocytes while a lower expression is
CC   observed in prostate, spleen, testis, ovary, small intestine,
CC   heart, liver, adrenal and thyroid glands.
CC   -!- PTM: PHOSPHORYLATED. THE NON-PHOSPHORYLATED FORM(S) EXHIBITED THE
CC   STRONGEST POLY(RC)-BINDING ACTIVITY.
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   -!- CAUTION: There is probably a frameshift error in the nucleotide
CC   sequence of Ref.3.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X78137; CAA55016.1; -.
DR   EMBL; U24223; AAA91317.1; -.
DR   EMBL; Z29505; CAA82631.1; ALT_FRAME.
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8647; PCBP1.
DR   GK; Q15365; -.
DR   MIM; 601209; -.
DR   GO; GO:0005737; C:cytoplasm; NAS.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003723; F:RNA binding; IDA.
DR   GO; GO:0003697; F:single-stranded DNA binding; IDA.
DR   GO; GO:0016071; P:mRNA metabolism; NAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Phosphorylation; Repeat.
FT   DOMAIN   13   75   KH 1.
FT   DOMAIN   97   162   KH 2.
FT   DOMAIN   279   343   KH 3.
FT   CONFLICT   205   205   V -> A (IN REF. 2 AND 3).
SQ   SEQUENCE   356 AA; 37526 MW; DC8547757GDC5104 CRC64;

Alignment Scores:
Pred. No.:   1
Score: 234.50
Percent Similarity: 41.51%   Length: 356
Best Local Similarity: 23.72%   Matches: 88
Query Match: 3.83%   Conservative: 66
DB: Mismatches: 138
                                           Indels: 79
                                           Gaps: 13

Pred. No.:   2.17e-07

```
Db      322 Ser------SerGlyArgGlnValThrIleThrGlySerAlaAlaSerIleSerLeuAla 339

Qy     1534 CAGGGACCGGATCTTTGGGAAACTGAAAGAGGAA 1566
            |||      ||||||||||||:::::||||       |||
Db      340 GlnTyrLeuIleAsnAlaArgLeuSerSerGlu 350


RESULT 15
PCB4_HUMAN
ID   PCB4_HUMAN          STANDARD;        PRT;       403 AA.
AC   P57723;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Poly(rC)-binding protein 4 (Alpha-CP4).
GN   PCBP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Single-stranded nucleic acid binding protein that binds
CC       preferentially to oligo dC (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 3 KH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
US-09-270-437D-6 (1-3412) x PCB1_HUMAN (1-356)

Qy    619 GGGGGACTTCTCAGGCCAGACAGATGATTCCGCTGCGGATCCTGTCCCACCCAG 678
          |||     |||          :::|||||||||:::|||      :::
Db      4 GlyValThrGlySerGlySerProLeuAsnValThrLeuThrIleArgLeuLeuMetHisGlyLys 23

Qy    679 TTTGTTGGTGCCATCATCGGAAAAGAGGAGGCTTGACCATAAAGAACATCACTAAGGACAGACC 738
          |||||:||||||||||||||||||||||:|||
Db     24 GluValGlySerIleIleGlyLysLysGlyGluSerValLysArgIleArgGluGlyGluSer 43

Qy    739 CAGTCCCGGCTAGATATCCATAGAAAAAGAAACTCGGAGCTGCAGAGAAGCCTGTCACC 798
          |||||||||||           |||||   ||:||||||||:|||
Db     44 GlyAlaArgIleAsnIle--------SerGluGlyAsnCysProGluArgIleIleThr 60

Qy    799 ATCCATGCCACCCCAGAGGGGACTTCTGAAGCATCCGCCATGAATTCTGAAATCATGCAG 858
          :::||||         |||||||      ::::::    ::::::
Db     61 LeuThrGlyProThrAsnAlaIlePheLysAlaPheAlaMetIleIleAspLysLysLeuGlu 80

Qy    859 AAAGAGGCAGATGAGACCAAACTA---------GCCGAAGAGATTCCTCTG 900
          ::::::     ::: |||  |||||||         |||||||||
Db     81 GluAspIleAsnSerSerMetThrAsnSerThrAlaAlaSerArgProProValThrLeu 100

Qy    901 AAAATCTTGGCACACAATGGCTTCGTGTTGGAGACTGATTGGAAAAGAGAGGCAGAAATTTG 960
          :::::::::|||       |||||||||||||||
Db    101 ArgLeuValValProAlaThrGlnCysGlySerLeuIleGlyLysGlyGlyCysLysIle 120

Qy    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAAATCTCATCTTTGCAGGATTTGAGC 1020
          |||       |||||||||||:::      ::::::      :::||||||
Db    121 LysGluIleArgGluSerThrGlyAlaGlnIleValGlnIleValAlaGly-----AspMetLeu 138

Qy   1021 ATATACAACCCGGAAGAACCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
          :::     |||||||||||||||||||||||||||||||||||
Db    139 ProAsnSerThrGluArgAlaIleThrIleAlaGlyValProGlnSerValThrGluCys 158

Qy   1081 GAGATAGAGATT-----------ATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATG 1128
          :::||||          :::::::     ::: |||
Db    159 ValLysGlnIleCysLeuValMetLeuGluThrLeuSerGlnSerProGlnGlyArgVal 178

Qy   1129 CTGGGTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATC 1188
          :::      :::          ||| ::::::
Db    179 MetThrIleProTyrGlnPro-------------MetProAlaSerSerProValIle 193

Qy   1189 TTTTCAACAGGACTTGTCCGTGCTATCTCCACCAGCAGGCTCCCCCGGAGCTCCCCCGCT 1248
          |||                ||||||
Db    194 CysAlaGlyAlaGlyGlnAspArgCysSerSerAspAlaValGly------------- 206

Qy   1249 GCCCCCTACCACCCCTTCACTACCCAC------------------TCCGGATACTTC 1287
          :::||||          |||||
Db    207 -------TyrProHisAlaThrHisAspLeuGluGlyProLeuGlyProProLeuAspAlaTyr-- 222

Qy   1288 TCCAGGCTGTACCCCCATCACCAGTTGGCCCG-----------------CAGGAGATT 1320
          |||||          ::: ||| ::::::
Db    223 --SerIleGlnGlyGlnHisThrIleSerProLeuAlaLeuAlaLysLeuAsnGlnVal 241

Qy   1321 ---------------------TTCCCGGATCATCAC----------- 1335
          :::                  |||||
Db    242 AlaArgGlnInSerHisPheAlaMetMetHisGlyThrGlyThrGlyPheAlaGlyGlyIleAsp 261

Qy   1336 ---TCTTATCCAGAG------------------------CAGGAGATT 1356
          |||                         |||||||
Db    262 SerSerSerProGluValLysGlyTyrTrpAlaSerLeuAspAlaSerThrGlnThrThr 281

Qy   1357 GTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCCATCATGCCGGAAGAAGGGGGCACAC 1416
          |||       |||          |||||||||||:::||||||
Db    282 HisGluLeuThrIleProAsnAsnLeuLysLeuGlyCysIleIleGlyArgGlnGlyAlaAsn 301

Qy   1417 ATCAAACAGCTGGCCGAGATTGCCCGGAGCCTCATCAAGATTGCC---CCTGCCGGAAGGC 1473
          |||       ::::::|||  ||||||||||:::||||||||||
Db    302 IleAsnGlyIleArgGlnMetSerGlyAlaGlnIleLysLysIleAlaAsnProValGluGly 321

Qy   1474 CCAGACGTCAGCGAAGAAGGGATGGTCATCATCACCGGGCCACCCGGAAGCCCAGTTCAAGGCC 1533
          |||       ||||||
```

```
CC
DR   EMBL; AF176330; AAG09241.1; -.
DR   EMBL; AK023993; BAB14761.1; -.
DR   EMBL; BC003008; AAH03008.1; -.
DR   EMBL; BC004153; AAH04153.1; -.
DR   Genew; HGNC:8652; PCBP4.
DR   GO; GO:0005634; C:nucleus; ISS.
DR   GO; GO:0003677; F:DNA binding; ISS.
DR   GO; GO:0003723; F:RNA binding; NAS.
DR   GO; GO:0016071; P:mRNA metabolism; ISS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       17     67       KH 1.
FT   DOMAIN      101    154       KH 2.
FT   DOMAIN      241    293       KH 3.
SQ   SEQUENCE   403 AA;  41481 MW;  3D99F762A9471265 CRC64;

Alignment Scores:
Pred.No.:                              403
Score:            3.47e-07
                  231.50               Matches:        89
Percent Similarity:    40.40%         Conservative:   54
Best Local Similarity: 25.14%         Mismatches:    106
Query Match:           3.78%          Indels:        105
DB:                    1              Gaps:           15

US-09-270-437D-6 (1-3412) x PCB4_HUMAN (1-403)

Qy   619 GGGGGCACTTCTCAGGCCAGACAG------ATTGATTCCCGCTGCGGATCCTGGTCCCC 672
               ||||||          ::|    |||||:::|||
Db     6 GlyGlyLeuGluGluProGluLeuSerIleThrLeuArgMetLeuMetHis 25

Qy   673 ACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAG 732
             |||::::|    ||||:::||||:||||:|| |||:|::|
Db    26 GlyLysGluValGlySerIleIleGlyLysLysGlyGluThrValLysArgIleArgGlu 45

Qy   733 CAGACCCAGTCCCGGGTAGATATCCATAGAAAGAAGAACTCTGGAGCTGCAGAGAAGCCT 792
               |||:||:::|||                    ::: |||:|:::
Db    46 GlnSerSerAlaArgIleThrIle---------SerGluGlySerCysProGluArg 61

Qy   793 GTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCCGCATGATTCTTGAAATC 852
             |||:|:|::|     ||||||:::|| |:||:::
Db    62 IleThr------ThrIleThrGlySerThrAlaAlaValPheHisAlaAlaValSerMet 78

Qy   853 ATGCAGAAAGAGCAGATGAGACCAAACTAGCC------------------- 885
             ::::|||    |||||||    ||
Db    79 IleAlaPheLysAspGluAspLeuCysAlaAlaProAlaAlaAsnGlyAlaAsnValSer 98

Qy   886 --GAAGAGATTCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTCGATTGGA 942
             |||:||    |||:::||||::::|||| |||:|||| |:| ||
Db    99 ArgProProValThrLeuArgLeuValIleIleProAlaSerGlnIleGlySerLeuIleGly 118

Qy   943 AAAGAAGGCAGAAATTGAAGAAAATTGAACATAACAGGGAACAAGATAACAATTCA 1002
             ||||||          :|:|| |:|::| |||:||||:|
Db   119 LysAlaGlyThrLysIleLysGluIleArgGluThrThrGlyAlaGlnInValAla 138

Qy  1003 TCTTTGCAGGATTTGAGCATCATATACACCCCGAAGCAATCACTGTGAAGGGCACAGT 1062
             |||||          :| |||||    ||:|:|||| ||||:::|||||
Db   139 Gly------AspLeuLeuProAsnSerThrGluArgValAlaValThrArgGlyValPro 156

Qy  1063 GAGGGCCTGTCCCAGTTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGGCCTTTGAAAAT 1122
             ::::|||
Db   157 AspAlaIle------------------------------------------- 159

Qy  1123 GATATGTCTGGCTGTTAACCAACAAGCAATCTGATCCCAGGGTTGAACCTCAGCGGCCACTT 1182
             |||||:|          :::|||
Db   160 ---IleLeuCysValArgGlnIleCysAlaValIle------ 170

Qy  1183 GGCATCTTTTCAACAAGGACTGTCCCGTCCTATCTCCACCAGCAGGGCCCCCGGAGCTCCC 1242

Db   171 ------------------LeuGluSerProProLysGly- 177
                             :::|||||||||   |||
Qy  1243 CCCGCTGCCCCCTACCACCCC------------TTCACTACCCACTCC 1278
                 |||||||||      |||  |||
Db   178 AlaThrIleProTyrHisProSerLeuGlyThrValLeuLeuSerAlaAsnGln 197

Qy  1279 GGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGC------------ 1317
                 |||::       |||:::||
Db   198 GlyPhe--------SerValGlnGlyValGlnTyrGlyAlaValThrProAlaLaGluVal 213

Qy  1318 ------------------CCGTTCCGGCCATCATCACTCTTATCCA 1344
                                          ||||
Db   214 ThrLysLeuGlnInLeuSerSerHisAlaValProPheAlaThrProSerValPro 233

Qy  1345 ------------GAGCAGGAGATTGTGAATCTCTTCATCCCAACC 1377
                           :::  |||
Db   234 GlyLeuAspProGlyThrGlnThrSerSerGlnGluPheLeu----ValProAsn 250

Qy  1378 CAGGCTGTGGGCCGCCATCATCGGGAAGAAGAAGGGGCCACACATCAAACAGTGCGGCGAGATTC 1437
                 :::||||       :::|||||      :::::
Db   251 AspLeuIleGlyCysValIleGlyLysArgGlyGlnLysIleSerGlyIleLeuArgGlnInMet 270

Qy  1438 GCCCGGAGCCTCTATCAAGATTGCC---CCTGCCGAAGGCCCAGCGTCAGCGGAAAGGATG 1494
                 :::|||||     ||||||      ||||:::|||
Db   271 SerGlyAlaHisIleLysIleGlyIleGlyIleGly------AlaGlyGluArgHis 288

Qy  1495 GTCATCATCACCCGGCCACCGGAAGCCCAGTTCAAGGGCCCAG 1536
                 |||  |||::|   :::  |||||
Db   289 ValThrIleThrGlySerProValSerIleAlaLeuAlaGln 302
```

Search completed: July 16, 2004, 10:45:29
Job time : 60.5 secs

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: July 16, 2004, 10:30:59 ; Search time 163.5 Seconds
                                      (without alignments)
                          13168.784 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 6121
Sequence: 1 ggcacgcggaggaggcgagga........aaccttgaaaatgttattt 3412

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09270437/runat_16072004_113049_13327/app_query.fasta_1.3591
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437_@CGN_1_1_283_@runat_16072004_113049_13327 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   SPTREMBL_25:*
         1:   sp_archea:*
         2:   sp_bacteria:*
         3:   sp_fungi:*
         4:   sp_human:*
         5:   sp_invertebrate:*
         6:   sp_mammal:*
         7:   sp_mhc:*
         8:   sp_organelle:*
         9:   sp_phage:*
        10:   sp_plant:*
        11:   sp_rodent:*
        12:   sp_virus:*
        13:   sp_vertebrate:*
        14:   sp_unclassified:*
        15:   sp_rvirus:*
        16:   sp_bacteriap:*
        17:   sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result   %   Query
No.   Score Match Length DB ID   Description
---- ----- ----- ------ -- ----   -----------
 1   2806.5 45.9   556   4   Q9Y6M1   Q9y6m1 homo sapien

	2	2358.5	38.5	545	11	Q7TQF9	Q7tqf9 mus musculu
	3	2053.5	33.5	593	13	O57526	O57526 xenopus lae
	4	2050	33.5	594	13	O73932	O73932 xenopus lae
	5	2033	33.2	576	13	O42254	O42254 gallus gall
	6	2023.5	33.1	577	11	Q8CGX0	Q8cgx0 rattus norv
	7	2021.5	33.0	577	11	O88477	O88477 mus musculu
	8	2015.5	32.9	577	4	Q9NZI8	Q9nzi8 homo sapien
	9	2014.5	32.9	577	11	Q8BRH1	Q8brh1 mus musculu
	10	2014.5	32.9	577	11	Q80US9	Q80us9 mus musculu
	11	1976	32.3	582	13	Q9PW80	Q9pw80 brachydanio
	12	1962.5	32.0	579	11	Q8C2J9	Q8c2j9 mus musculu
	13	1957.5	31.9	579	11	Q9CPN8	Q9cpn8 mus musculu
	14	1950.5	31.9	579	4	O00425	O00425 homo sapien
	15	873.5	14.3	580	5	Q8IGK4	Q8igk4 drosophila
	16	869	14.2	573	5	Q8IR99	Q8ir99 drosophila
	17	866.5	14.2	566	5	Q9VZ69	Q9vz69 drosophila
	18	493.5	8.1	828	5	Q21605	Q21605 caenorhabdi
	19	337	5.5	100	4	Q86VB1	Q86vb1 homo sapien
	20	279.5	4.6	568	5	Q9LXF5	Q9lxf5 arabidopsis
	21	277.5	4.5	398	3	O74919	O74919 schizosacch
	22	269.5	4.4	833	10	Q9FMF0	Q9fmf0 arabidopsis
	23	269	4.4	621	10	Q9C553	Q9c553 arabidopsis
	24	267.5	4.4	762	10	Q8LNT9	Q8lnt9 oryza sativ
	25	265.5	4.3	680	5	P91393	P91393 caenorhabdi
	26	263	4.3	641	5	Q9BLA0	Q9bla0 caenorhabdi
	27	262	4.3	351	4	Q8N9K6	Q8n9k6 homo sapien
	28	260.5	4.3	169	11	Q7TP50	Q7tp50 rattus norv
	29	259	4.2	557	5	Q23487	Q23487 caenorhabdi
	30	257.5	4.2	313	4	Q96EP6	Q96ep6 homo sapien
	31	254.5	4.2	589	5	Q17935	Q17935 caenorhabdi
	32	254.5	4.2	611	5	Q17936	Q17936 caenorhabdi
	33	254	4.1	836	10	Q84ZW9	Q84zw9 arabidopsis
	34	252	4.1	510	5	Q72145	Q72145 caenorhabdi
	35	249.5	4.1	869	10	Q84ZX0	Q84zx0 arabidopsis
	36	249	4.1	370	11	Q8BBB0	Q8bbb0 mus musculu
	37	248.5	4.1	371	11	Q8C544	Q8c544 mus musculu
	38	246.5	4.0	644	10	Q8LDV1	Q8ldv1 arabidopsis
	39	246.5	4.0	644	10	Q9FNK3	Q9fnk3 arabidopsis
	40	245.5	4.0	318	13	Q7ZVK5	Q7zvk5 brachydanio
	41	239.5	3.9	640	10	Q9ASX3	Q9asx3 arabidopsis
	42	239	3.9	542	10	Q8S7G1	Q8s7g1 oryza sativ
	43	238.5	3.9	532	10	Q8HOY7	Q8h0y7 arabidopsis
	44	237.5	3.9	653	13	Q72XS1	Q72xs1 xenopus lae
	45	236.5	3.9	403	11	Q8BP63	Q8bp63 mus musculu

ALIGNMENTS

RESULT 1
Q9Y6M1
ID   Q9Y6M1   PRELIMINARY;   PRT;   556 AA.
AC   Q9Y6M1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hepatocellular carcinoma autoantigen.
GN   P62.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207072; PubMed=10190901;
RA   Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT   "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT   in human hepatocellular carcinoma.";
RL   J. Exp. Med. 189:1101-1110(1999).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF067352; AAD31596.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.

```
DR   GO; GO:0009386; P:translational attenuation; TAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR005504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   556 AA;  61842 MW;  1EDEFB100443DDC4 CRC64;
```

Alignment Scores:
```
Pred. No.:
Score:                 2806.50          Length:          556
Percent Similarity:    92.82%           Matches:         556
Best Local Similarity: 92.82%           Conservative:    0
Query Match:           45.85%           Mismatches:      0
DB:                                     Indels:          43
                                        Gaps:            1
```

US-09-270-437D-6 (1-3412) x Q9Y6M1 (1-556)

```
Qy    70 ATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGGCCGCGACCTCCGG 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetMetAsnLysLeuTyrIleGlyAsnLeuSerProAlaValThrAlaAspAspLeuArg 20

Qy   130 CAGCTCTTTGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTAC 189
          ||||||||||||||||||||||||
Db    21 GlnLeuPheGlyAspArgArgLysLeuProLeuAlaGlyAlaGlyGlnValLeuLeuLysSerGlyTyr 40

Qy   190 GCCTTCGTGACTACCCCGACCAGAACTGGGCCATCGGCCATCGAGACCCTCTCGGGT 249
          ||||||||||||||||||||||||||||||||||||
Db    41 AlaPheValAspTyrProAspGlnAsnTrpAlaIleAlaIleArgAlaIleGluThrLeuSerGly 60

Qy   250 AAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGG 309
          |||||||||||||||||||||||||
Db    61 LysValValGluLeuHisGlyLysIleMetGluValAspTyrSerValSerLysLysLeuArg 80

Qy   310 AGCAGGAAAATTCAGATTCTGAAACATCCCTCCTCACCTGCCAGTGGGAGGTGTTGGATGGA 369
          ||||||||||||
Db    81 SerArgLysIleGlnIleLeuArgAsnIleProProHisLeuGlnGlnTrpGluValLeuAspIle 100

Qy   370 CTTTTGGCTCAATATGGACAGTGGAGAATGTGAACAACAGTCAACACAGACAGGAAACC 429
          |||||||||||||||||||||||||||
Db   101 LeuLeuAlaGlnTyrGlyGlyThrValGluAsnValGluGlnAsnThrAspThrGluThr 120

Qy   430 GCCGTTGTCAACGTTCACATATGCAACAGAGAAGAGCAAAAATAGCCATGAGGAGAAGCTA 489
          |||||||||||||||||||||||||||||
Db   121 AlaValValAsnValThrTyrAlaThrArgArgGluGluGlnAlaMetGluAlaMetGluGluLysGluLeu 140

Qy   490 AGCGGGCATCAGTTGGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTG 549
          |||||||||||||||||||||
Db   141 SerGlyHisGlnLeuGluAsnTyrSerPheLysIleSerTyrIleProAspGluGluVal 160

Qy   550 AGCTCCCCTTCGCCCCCTCAGCGAGCCCGTGGGGGACCCACTCTTCCCGGGAGCAAGGC 609
          |||||||||||||||||||||
Db   161 SerSerProSerProProGlnArgAlaArgGlyAlaGlnIleArgGlyAlaGlnLeuSerSerArgGluGlyGly 180

Qy   610 CACGCCCCTGGGGGCACTTCTCAGGGCACAGATTCGATTCCCGGGCGGCGATCCTGGTC 669
          |||||||||||||
Db   181 HisAlaProGlyGlyThrSerGlnIleLeuAlaArgGlnIleIleAspPheProLeuArgIleLeuVal 200

Qy   670 CCCACCCAGTTTGTTGGTGCCATCATCGCGGAAAAGGAGGCCTTGACCATAAAGAACATCACT 729
          ||||||||||||||||
Db   201 ProThrGlnPheValGlyAlaIleIleGluGlyLysLeuThrIleLeuLysAsnIleThr 220

Qy   730 AAGCAGACCCAGTCCGGGTAGAATATCCATAGAAAAGAAGAACCTGAGCTCAGAGAAG 789
          |||||||||||
Db   221 LysGlnThrGlnSerArgValAspIleHisArgGlnIleHisSerGlyLysAsnSerGlyAlaAlaAlaLys 240
```

```
Qy   850 ATCATGCAGAAAGAGGCAGATAGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTG 909
          |||||||||||||||||||||||||||||||||||||||||||
Db   261 IleMetGlnLysGluAlaAspArgAspGlnThrLeuAlaGluGluIleProLeuLysIleLeu 280

Qy   910 GCAACACAATGGCTTGGTTGGAGACTGATTGAAAGAAGGCAGAAATTTGAAGAAATT 969
          ||||||||||||||||||||||||||||||
Db   281 AlaHisAsnGlyLeuValGlyArgLeuIleGluArgArgGlnLysLeuLysLysIle 300

Qy   970 GAACATGAAACAGGGACCAAGATAACAATCTCATCTTCAGATTTGAGCATATACAAC 1029
          |||||||||||||||||||||||||||||||||
Db   301 GluHisGluThrGlyThrLysIleThrIleSerSerSerLeuGlnAspLeuSerIleTyrAsn 320

Qy  1030 CCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGCCAGTGCTGAGATAGAG 1089
          |||||||||||||||||||||||||||||||||
Db   321 ProGluArgThrIleThrValLysGlyThrValGluProAlaSerAlaGluIleGlu 340

Qy  1090 ATTATGAAGAAGCTGCGTGAGCGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAGCC 1149
          ||||||||||||||||||||||||||||||
Db   341 IleMetLysLysLeuArgGluArgLeuAlaPheGluAsnAspMetLeuAlaValAsn-------- 357

Qy  1150 AATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCGTG 1209
Db   357 -------------- 357

Qy  1210 CTATCTCCACCAGCAGGCCCCCGGCGGAGCTCCCCGCCCTGCCCTCCCACCACCCCTTCACT 1269
Db   357 -------------- 357

Qy  1270 ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCGTTCCCGCAT 1329
          |||||||||||||||||||||||||
Db   358 ThrHisSerGlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProPheProHis 377

Qy  1330 CATCACTCTTATCCAGAGCAGGAGATTGTGAATCCTCATCCCAACCCAGGCTGTGGGC 1389
          ||||||||||||||||||||||||||||
Db   378 HisHisSerTyrProGluGlnGluIleValAsnLeuPheIleProThrGlnAlaValGly 397

Qy  1390 GCCATCACTCGGGAAGAAGGGGCCACACATCAACAGCTGCGGGAGATTGCGCGGAGCCTCT 1449
          |||||||||||
Db   398 AlaIleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPheAlaGlyAlaSer 417

Qy  1450 ATCAAGATTGCCCTGCCGAAGGCCCCAGAGCTCAGCGAAAGGATGGTCATCATCACCGGG 1509
          ||||||||||||||||||||||||||||
Db   418 IleLysIleAlaProAlaGluGluProAspValSerGluArgMetValIleIleIleThrGly 437

Qy  1510 CCACCCGAAGCCCAGTTCAAGGCCCAGGACGATCTTTGGGAAACTGAAAGAGGAAAAC 1569
          ||||||||||||||||||||||||||||||
Db   438 ProProGluAlaGlnPheLysAlaGlnAspAspLeuTrpGluLeuGluLysGluGluAsn 457

Qy  1570 TTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGGGCCATATCAGAGTGCCCTCTTCCACA 1629
          |||||||||||||||||||||
Db   458 PhePheAsnProLysLysLysLeuValLysLeuGluLysAlaHisIleArgValProSerSerThr 477

Qy  1630 GCTGGCCGGGTGATTGGCAAAGGTGGCAAAGCCGTGAACAACTGCAGAACTTAACCAGT 1689
          |||||||||||||||||||||
Db   478 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrSer 497

Qy  1690 GCAGAAGTCATCGGTGCTGCTCGTCGACCAAAACCAGATGAAAATGAGGAAGTGGATGTCAGA 1749
          |||||||||||||
Db   498 AlaGluValIleValProArgAspGlnThrProAspGluAsnGluGlnValIleValArg 517

Qy  1750 ATTATCGGGCCACTTCTTTGCTAGCCAGAACTGCACACGCCAAGATCAGGGAAATTGTACAA 1809
          |||||||||||
Db   518 IleIleGlyHisPhePheAlaSerGlnThrAlaGlnArgLysIleArgGluIleValGln 537

Qy  1810 CAGGTGAAGCAGCAGAGACAGAAATACCTTCAGGAGTCGCCTCACAGCCCAGCAAG 1866
          ||||||||||||||||
Db   538 GlnValLysGlnGlnArgGlnLysTyrProGlnGluValAlaAlaSerGlnArgSerLys 556
```

RESULT 2

```
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC054552; AAH54552.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   545 AA;  59679 MW;  88DD35FF30DF3091 CRC64;
```

Alignment Scores:

```
Pred. No.:              7.16s-167          Length:          545
Score:                  2358.50            Matches:         456
Percent Similarity:     95.71%             Conservative:    13
Best Local Similarity:  93.06%             Mismatches:      14
Query Match:            38.53%             Indels:          7
                                           Gaps:            1
```

US-09-270-437D-6 (1-3412) x Q7TQF9 (1-545)

```
Qy   73  ATGAACAAGCTTTACATCGGAACTGAGCCCCGTCACCGGCCGACGACCTCCGGCAG  132
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MetAsnLysLeuTyrIleGlyThrGluProValThrAlaAspAspLeuArgGln    20

Qy  133  CTCTTTGGGGACAGGAAGTGCCCCTGCGGGACAGGTCCTGAAGTCCGGCTACGCC  192
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21  LeuPheGlyAspArgLysLysLeuProLeuArgGlyGlnValLeuLysSerGlyTyrAla  40

Qy  193  TTCGTGGACTACCCCGACCAGAAGTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAA  252
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41  PheValAspTyrProAspGlnLysTrpAlaIleArgAlaIleGluThrLeuSerGlyLys  60

Qy  253  GTGGAATTGCATGGGAAAAATCATGAAAGTTGATTACTCAGTCTCTAAAAAGCTAAGAGC  312
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ValGluLeuHisGlyLysAsnHisGluSerLeuPheLeuSerValSerLysLysLeuArgSer  80

Qy  313  AGGAAAAATTCAGATCGAAATCCCTCCTCACCTGCAGTGGAGAGGTGTTGGATGGACTT  372
         |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81  ArgArgIleGlnIleGluProHisLeuGlnTrpArgGlyValLeuAspGlyLeu  100

Qy  373  TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACACAGAAACCCGCC  432
         |:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101  LeuAlaGluTyrGlyThrValGluAsnValGluGlnValAsnThrAspThrGluThrAla  120

Qy  433  GTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAATAGCCATGAGAAGCTAAGC  492
         ||| |||||||| | |||||| | :::|||
Db  121  ValValAsnValThrTyrMetThrArgGluValAlaLysLeuAlaIleGluLysLeuSer  140

Qy  493  GGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC  552
         |||||:::||||:|||||||||||||||||||||||||||||||||||||
Db  141  GlyHisGlnPheGluAspTyrSerPheLysIleProThrSerProAspGluValSer  160

Qy  553  TCCCTTGCCCCTCAGCGAGCCAGCGTGGGACCACTCT···CCCGGGAGCAAGGCCAC  612
         |||||||||||||||||||||||||||||||------------
Db  161  SerProSerProProHisArgAla-----------ArgGluGlnGlyHis  173

Qy  613  GCCCCTGGGGCACTTCTCAGGCCAGACAGATGATTTCCCGCTGCGGATCCTGGTCCCC  672
         ||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db  174  GlyProGlySerSerGlnAlaArgGlnIleAspPheProLeuArgIleLeuValPro  193

Qy  673  ACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAG  732
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194  ThrGlnPheValGlyAlaIleIleGlyLysGluGlyLeuThrIleLysAsnIleThrLys  213

Qy  733  CAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGCCT  792
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  214  GlnThrGlnSerArgValAspIleHisArgLysGluAsnSerGlyAlaAlaGluLysPro  233

Qy  793  GTCACCATCCATGCCAGGGGACTTCTGAAGCATGCCCATGATCCTGAAAATC  852
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  234  ValThrIleHisAlaThrProGluGluGlyThrSerGluAlaCysArgMetIleLeuGluIle  253

Qy  853  ATGCAGAAAGAGGCAGATGAGACCAACTAGCCAAGAGATTCCTCGAAAATCTTGGCA  912
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254  MetGlnLysGluAlaAspGluThrLysLeuLeuAlaGluValProLeuLysIleLeuGluAla  273

Qy  913  CACAATGGCTTGGTTGGAAGACTGATGGAAAGAAGGCAGAAATTTGAGAAAAATTGAA  972
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274  HisAsnGlyPheValGlyArgLeuIleGluLysLysGluGluIleLysLysLysLeuGlu  293

Qy  973  CATGAAAACAAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATAACAACCCG  1032
         ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db  294  HisGluThrLysLysGlnAspAsnAsnLeuIlePheAlaGlyPheGluHisAsnAsnPro  313

Qy 1033  GAAAGAACCATCACTGTGGAAGGGCACAGTTGAGGCCTGAGCCTGCCAGTGCTGAGATAGAGATT  1092
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314  GluArgThrIleThrValGluGlyHisSerLeuSerAlaLeuAsnAlaLysCysAlaAsnAlaGlyIleIleGlnIle  333

Qy 1093  ATGAAGAAGCTGCGTGAGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAAT  1152
         ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db  334  MetLysLysLeuArgGluProLeuLysMetIleCysTrpLeuLeuThrAsnGlnProAsn  353

Qy 1153  CTGATCCCAGGGTTGAACCTCAGCGACTCAGCATCTTTCAACAGACTGTCCGTGCTA  1212
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354  LeuIleProGlyLeuAsnLeuSerAspSerAlaSerPheAsnArgLeuSerValLeu  373

Qy 1213  TCTCCACCAGCAGGCCCCGAGCTCCCCGGCAGGTCCGGAAGTCCGGCTACACC  1272
         |||||||||||||||||||||:::|||||||||||||||||
Db  374  SerProProAlaGlyProGluLeuProGlyArgSerGlyValProProTyrHisProAlaThr  393

Qy 1273  CACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCGTTCCCGCATCAT  1332
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  394  HisSerGlyTyrPheSerSerLeuTyrProHisHisGlnPheGlyProProHisHis  413

Qy 1333  CACTCTTATCCAGACAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCC  1392
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414  HisSerTyrProGlnArgArgAspLeuGluSerLeuProSerLeuTyrSerGlnLysAspAlaValGlyAla  433

Qy 1393  ATCATCGGGAAGAAGGGGGCACACATCAAACAGTCGGCGGAGCCTCATC  1452
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  434  IleIleGlyLysLysGlyAlaHisIleLysGlnSerAlaGluProHisSerAlaSerIle  453

Qy 1453  AAGATTGCCCCTGCGAAGGCCCAGAATGTGGAACAAGTCAACACAGACACAGAAACCCGCC  1512
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  454  LysIleAlaProAlaLysAlaGlnAsnValGluGlnValAsnThrAspThrGluThrPro  473
```

```
Qy    1513  CCGAAGCCAGTTCAAGGCCCAGGGACGG 1542
                 ||||||||||||||||||||||
Db     474  ProGluAlaGlnPheLysValSerAlaLys 483
```

RESULT 3
O57526

```
ID   O57526               PRELIMINARY;     PRT;     593 AA.
AC   O57526;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   KH domain-containing transcription factor B3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92249652; PubMed=1577195;
RA   Pfaff S.L., Taylor W.L.;
RT   "Characterization of a Xenopus oocyte factor that binds to a
RT   developmentally regulated cis-element in the TFIIIA gene.";
RL   Dev. Biol. 151:306-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Griffin D., Taylor W.L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF042353; AAB97457.1; -.
DR   EMBL; AF064633; AAC18597.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE 593 AA;  65385 MW;  5A5AB4B4A1D55DF7 CRC64;
```

Alignment Scores:

```
Pred. No.:                        Length:         593
Score:            3.8e-144        Matches:         403
Percent Similarity:  79.67%       Conservative:     75
Best Local Similarity: 67.17%     Mismatches:       87
Query Match:      33.55%          Indels:           35
DB:               593 AA;         Gaps:              5
```

US-09-270-437D-6 (1-3412) x O57526 (1-593)

```
Qy     73  ATGAACAAGCTTTACATCGGGAACCTGGGAACCTGAGCCCGTCACCGCCGCGACGACCTCCGGCAG 132
               |||||||||||||||||||||  ||||||  |||:::           |||||
Db      1  MetAsnLysLeuTyrIleGlyAsnLeuGlyThrLeuSerProSerProLeuAsnValSerProLeuAsnValSerProProAspLeuGluSer 20

Qy     61  ValGluLeuHisGlyLysValIleGluValAlaGluHisSerValProLysArgGlnArgSer 80
               |||||||||||||||||||||:::::||||
Db     313  AGGAAAATTACAGATTCCAAACATCCTCTCCACCTGCAGTGGGAGGTGTTGGATGGACTT 372

Db      81  ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
               |||:::|||||||||||||||:::|||||||||
Db     373  TTGGCTCCAATATGGGGACAGTGGAGAATGTGGAACAGTGCAACACAGACACAGAAACCGCC 432

Db     101  LeuAlaGlnPheGlyThrValGluAsnCysGluGlnValAlaSnThrAspSerGluThrAla 120
               ||||||||||||||||  ||:::||||||||
Qy     433  GTTGTCACGTCACATGCACAAGAGAGAGAAGAGCCAAAATATGCCATGGAGAAGCTAAGC 492

Db     121  ValValAsnValThrTyrAlaAsnlysGluHisAlaArgGlnGlyLeuGluLysLeuAsn 140
               |||||||||||||||||:::||||||||||
Qy     493  GGGCATCAGTTGAGAACTACTCCTTCAAGATTCTACATCCCGATGAAGAGGTGAGC 552

Db     141  GlyTyrGlnLeuGluAsnTyrSerLeuLysValThrTyrIleProAspGluMetAlaThr 160
               :::||||||||||||||||||||||||:::
Qy     553  TCCCCTTCGCCCCCTCCAGCGAGCCCAGCGT 582

Db     161  ProGlnSerProSerGlnInLeuGlnProGlnGlnHisProGlnGlnHisProGlnGlyArgArg 180
               ||||||||||||||||||||  ||:::
Qy     583  GGGGACCACTCTTCCCGGGACCAAGGCCACGCCCTGGGGCACTTCTCAGGCC 636

Db     181  GlyPheGlyGlnArgGlyProAlaArgGlnGly--SerProGlyAlaAlaAlaArgPro 199
               |||||||||||||||||||||  ::::||||||
Qy     637  AGA---CAGATTGATTTCCCGCTGCGATCCTGGTCCCCACCCAGTTGTTGGTGCCATC 693

Db     200  LysProGlnSerGluValProLeuArgMetLeuValProThrMetGlnPheValGlyAlaIle 219
               |||:::|||||||||||||||||||||||||||||||||
Qy     694  ATCGGAAGGAGGGCTTGACCATAAGAACATCACTAAGCAGCACCAGTCCGGGTAGAT 753

Db     220  IleGlyLysGlyGlyAlaThrIleArgAsnIleThrGlnSerGlnSerIleAsp 239
               ||||||||||||||||||||||::::|||||||||::::
Qy     754  ATCCATAGAGAAAGAGAACTCTGGAGTGCGAGAGAAGCCTGTCCATGCCACCCCA 813

Db     240  IleHisArgLysGluAsnAlaGlyAlaAlaAlaGluLysProIleThrIleHisSerThrPro 259
               |||||||||||||||||||||||||||||||:::||||||||
Qy     814  GAGGGGACTTCGAAGCATGCCCATGATTCTTGAAATCATGCGAGAAGAGGCCAGATGAG 873

Db     260  GluGlyCysSerAlaAlaCysLysIleIleMetGluIleMetGlnGlnAlaGlnGlnAsp 279
               ||||||||||||||||||||||||||||||||||:::
Qy     874  ACCAAACTAGCCGAAGAGATTCCTGAAAATCTTGCCACACAATGGCTTGGTTGGAAGA 933

Db     280  ThrLysPheThrGluIleLeuProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 299
               |||||||||||||||||||||||||||||||||||
Qy     934  CTGATTGGAAAGAAGCAGAAATTGAAGAAATTGAACATGAAACAGGGACCAAGATA 993

Db     300  LeuIleGlyLysGluGluProLeuArgAsnLeuLysIleLysGluGlnLysAspThrLysIle 319
               ||||||||||||||||||||||||||||||||||||
Qy     994  ACAATCTCATCTTGCAGGATTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAG 1053

Db     320  ThrIleSerProLeuGlnAspLeuThrLeuTyrAsnProGlyArgThrIleThrValLys 339
               |||||||||::::||||||  |||||||::|||
Qy     1054 GGCACAGTTGAGGCCTGCCAGTGCTGAGATAGAGATTATGAAGAAGTGCGTGAAGGCC 1113

Db     340  GlySerIleGluThrCysAlaLysAlaGluLeuGluValMetLysIleLysArgGluSer 359
               |||||||||||||||||||||||||||||||||
Qy     1114 TTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCCAATCGATCCCAGGGTTGAACCTC 1173

Db     360  TyrGluAsnAspIleAlaMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeu 379
               :::||||||||||||||||||||||||||||||||
Qy     1174 AGGGCACTTGGCATCTTTTCAACAGGACTGTCGTGCTATCTCCACAGGCCCCGC 1233

Db     380  AsnAlaLeuGluPhe------------ProProSerSerGly 391
               :::||||||||||:::                ||||||
Qy     1234 GGAGCTCCCCGCTGCCCCCTACACCCCTACTACCCGATACTTCTCCAGC 1293

Db     392  MetProProSerAlaGlyValSerProThrArgSerAlaSerTyr------ 408
               ::::|||||||||||||||||||||||::||------
Qy     1294 CTTGTACCCCCATCACCAGTTTGGCCCGTTCCCGCCATCATCACTCTTATCCAGACAGGAGAG 1353
```

Alignment Scores:
Pred. No.:
Score: 7.07e-144
                            2050.00
Percent Similarity: 79.24% Length: 594
Best Local Similarity: 66.72% Matches: 405
Query Match: 33.49% Conservative: 76
DB: 13   Mismatches: 92
                                 Indels: 34
                                 Gaps: 6

US-09-270-437D-6 (1-3412) x O73932 (1-594)

```
Qy   73 ATGAACAAGCTTACATGGGAACCTGAGCCCCGAGACCTCCGGCAG 132
        |||||||||:|||:|||||||||||||||||||||||||||||||
Db    1 MetAsnLysLeuTyrIleGlyAsnLeuSerProThrAspLeuGluSer 20

Qy  133 CTCTTGGGACAGGAGTGCCCTGGCGGACAGGTCCTGCTGAAGAGTCGGCTACGGC 192
        ||||||:||::|||||||||||||||||||||||||||:::
Db   21 LeuPheLysGluSerLysIleProPheThrGlyGlnPheLeuValLysSerGlyTyrTyrAla 40

Qy  193 TTCGTGGACTACCCGACCAGAACTGGGCCATCGGCCATCGAGACCCTCCGGGTAAA 252
        |||||||||||||||||||||||:::::|||||||:::|||||
Db   41 PheValAspCysProAspGluThrTrpAlaMetIleGlyAlaIleAspThrLeuSerGlyLys 60

Qy  253 GTGGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
        ||||||||||||||||:::::|||||||||||:||:|||||||||||||||
Db   61 ValGluLeuHisGlyLysValIleGluValIleThrGlnSerValProLysGlnArgSer 80

Qy  313 AGGAAAATTCAGATTGAAACATCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTT 372
        ||||||:::||||||||||:||||||||||:|||||||||||||:::
Db   81 ArgLysLeuGlnIleGluThrSerSerSerProHisSerLeuGlnTrpGluValLeuAspGlySerLeu 100

Qy  373 TTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAACCGCC 432
        ||||:::|||||||||||||||||||||:|||:|||||:::
Db  101 LeuAlaGlnTyrGlyThrValGluAsnValGluGlnValAsnThrGluArgAsnArgAla 120

Qy  433 GTTGTCAACGTCACATATCCAACAAGAGAGAGAGAGAGATAGCCATGGAGAGCTAAGC 492
        |||||:::|||||:|||||||||||:||||||||||||||||||
Db  121 ValValAsnValThrTyrAlaAsnLysGluLeuHisAlaArgGlnGlyLeuGluLysLeuAsn 140

Qy  493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATCGAAGAGGTGAGC 552
        ||::||:|||||||:::|||||||||||||:::||||||||||
Db  141 GlyTyrGlnLeuGluAsnTyrSerLeuLysValThrTyrIleProAspAspMetAlaThr 160

Qy  553 TCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGGAGCAAGGCCAC 612
        |||||:|||||||||||||||||||||||||:::
Db  161 ProGlnAlaProSerGlnGlnGlnLeuGlnGlnLeuGlnGlnHisGlyGlnGlnHisGlyArg 180

Qy  613 ------------------------GCCCTGGGGCACTTCTCCAGGCG 636
                                ::::|||||
Db  181 ArgGlyPheGlyGlnArgGlyProAlaArgGlyLyaSerProGlyAlaAlaArgGlnPro 200

Qy  637 AGA---CAGATTGATTTCCCGTGCGGATCCTGGTCCACCCAGTTGTTGTGGTGCCATC 693
        :::    ::::|||||||||:||||||||||||
Db  201 LysProGlnThrGluLeuArgMetLeuValProLeuThrHisGlnInPheValGlyAlaIle 220

Qy  694 ATCGGAAAGGAGGGCCTTGACCATAAAGAACATCACTAAGACCAGTCCCAGTTCCCGGGTAGAT 753
        ||::|||:|||||||:||||:|||:|||||||||||||||||
Db  221 IleGlyLysGluGlyGlyAlaThrIleArgAsnIleThrLysGlnSerGlyAlaArgIleGlnLeuIleAsp 240

Qy  754 ATCCATAGAAAGAGAACTCTGGAGTCGAGAGAGCCCTGTCACCATGCCACCCCA 813
        |||||:::::|||||||||||||||||||||||||||||||||
Db  241 IleHisArgArgLysLysGlnAlaAsnAlaGlyAlaAlaGluLysProIleGlnSerProThrPro 260

Qy  814 GAGGGGACTTCTCAAGCATGCCCATGATTCTTGAAATCATGCAGAAAGAGAGGAGCAGATGAG 873
        ||:::|||||:|||||||||||||||:|||||||||||:||
Db  261 GluGlyCysSerAlaAlaCysLysLysIleLeuIleMetGluIleMetLysLysGluGlnGlnAsp 280

Qy  874 ACCAAACTAGCCGGAGAGGATTCCTGGAAAATCTTGCCACACAATGGCTTGGTTGGAGAA 933
        |||||::::|||||||||||||:|||||||||
Db  281 ThrLysPheThrGluIleLeuProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 300

Qy  934 CTGATTGAAAGAGAAGGCCAGAAATTTGAAGAAAATTGAACATGAAACAGGACCAAGATA 993
        |||||||||||:::||||||
Db
```

```
Db  409 ----------------------------ProProPheGlyGlnGlnProGluSerGlu 418

Qy 1354 ATTGTGAATCTCTTCATCCAACCCAGGCTGTGGGCGCCATCATCCCGGAAGAAGGGGGCA 1413
        ||||:::|||||||||:||||||||||||||||||||||||||||||||||||||
Db  419 ThrValHisLeuPheIleProAlaLeuAlaValGlyAlaIleIleGlyLysLysGlnGlyGln 438

Qy 1414 CACATCAAAACAGCTGGCGAGAGTTCGCCGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGC 1473
        ||||||:::|||||||:::|||||||||||||||||||||||||||||||
Db  439 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerArgPheAlaSerIleAlaProAlaGluGly 458

Qy 1474 CCAGAACGTCAGCGAAAGGATGGTCATCATCCACCCGGCGCCAAGCCCAGTTCAAGGCC 1533
        |||||||||||||:::|||||:::|||||||||||||||||||||||||||||||||
Db  459 ProAspAlaLysLeuArgMetValIleIleThrGlyProProGluAlaGlnPheLysAla 478

Qy 1534 CAGGGACGGATCTTTGGGAAACTGAAAGAGAGGAAAACTTCTTAACCCCAAAGAAGAAGTG 1593
        |||||||||||||||||||||||||:::|||||||||||||:::||||||||||||||||
Db  479 GlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGlyProLysLysGluGluVal 498

Qy 1594 AAGCTGGAAGGCCATATCAGAGTGCCCTCTTCACAGCTGCCGGGTGATTGGCAAAGGT 1653
        ||||||||||||:::|||||::|||||||||||||||||||||||||||||||||
Db  499 LysLeuGluAlaHisIleLysValProSerTyrArgLeuProGlyAspTrpLysGlyGly 518

Qy 1654 GGCCAGAACCTGAACGAACTGCAGAACTAACCAGTGCAGAAGTCATCGTGCCTCGTGAC 1713
        |||||||||||||:::|||||||||||:::|||||||||:||||||||||||||||
Db  519 GlyGlnLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluValValProArgAsp 538

Qy 1714 CAAACGCCAGATGAAAATGAGGAGGATGTCAGAATTATCGGGCACTCTTTGCTAGC 1773
        ||:::|||||||||||||:||||||||||:::||||||||:|||||||||||||
Db  539 GlnThrProAspGluAsnGluGluAspValArgIleIleGlyHisSerLeuLeuAlaSer 558

Qy 1774 CAGACTGCACAGCGCCAAGATCAGGGAAATTCTACACAGGTGAAGCAGCAGGAGCAGAAA 1833
        |||||||||||||:::|||||:::|||||
Db  559 GlnLeuAlaGlnArgLysIleGlnGluIleLeuGluLeuAlaGlyArgGlnGlnGln 578
```

RESULT 4
O73932

```
ID   O73932      PRELIMINARY;   PRT;   594 AA.
AC   O73932;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   VG1 RNA binding protein variant D.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98228351; PubMed=9560341;
RA   Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
RT   "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT   localization in vertebrates.";
RL   Curr. Biol. 8:489-496(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF064634; AAC18598.1; -.
DR   EMBL; AF055923; AAC41285.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR005004; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   594 AA;   65643 MW;   54CEA7BFF0855DD6 CRC64;
```

```
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97220007; PubMed=9121465;
RA   Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT   "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL   Mol. Cell. Biol. 17:2158-2165(1997).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF026527; AAB82295.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR005004; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00360; RRM; 2.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS00084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   576 AA;  63271 MW;  01AAF2D1D81C8811 CRC64;

Alignment Scores:
Pred. No.:                  1.29e-142      Length:            576
Score:                      2033.00        Matches:           402
Percent Similarity:         78.94%         Conservative:      74
Best Local Similarity:      66.67%         Mismatches:        95
Query Match:                33.21%         Indels:            32
DB:                         13             Gaps:              6
```

```
US-09-270-437D-6 (1-3412) x O42254 (1-576)

Qy     73  ATGAACAAGCTTACATCGGGAACTGAGACCCGCGTCACCGCCGAGACCTCCGGCAG  132
               :::|||||||||||||| |||||||||| |::::     :||||:|||||
Db      1  MetAsnLysLeuTyrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluGluLys  20

Qy    133  CTCTTTGGGACCAGGAAGCTGCCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
               ::: ||| ||| |||||||||||||||||||||:: ||||||||||:|||||||||
Db     21  ValPheAsnAspHisLysIleSerPheSerGlyGlnPheLeuValLysSerGlyTyrAla  40

Qy    193  TTCGTGGACTACCCCGACCAGAGCTGGGCCATCGGCCATCGAGACCCTCCGGGTAAA  252
               ||||||:::||||||||||||||| ||::|||||||||||||||:::|||||
Db     41  PheValAspCysProAspGlnSerTrpAlaMetLysAlaIleGluThrPheSerGlyLys  60

Qy    253  GTGGAATTGCATGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC  312
               |||||||||| || ||||||| ::||||:||||||:::|||||||||||||
Db     61  ValGluLeuHisGlyLysSerGlnLeuLeuGluHisSerValProLysGlnArgSer  80

Qy    313  AGGAAAATTCAGATTCGAAACTCCCTCCTCCACCTGCAGTGGCGAGGTGTTGGATGGACTT  372
               |||||||| ||||| |||||:::|||||||||||||||||||||:::||||||||
Db     81  ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpLeuValLeuAspGlyLeu  100

Qy    373  TTGGCTCAATATGGACAGTGGACAGATCAGGAGAATGTGAACAAGTCAACACAGAACCGCC  432
               ||||||||| |||||||||| ||||||:::||||| |||:::|||||||||||||
Db    101  LeuAlaGlnTyrGlyTyrValGluAsnCysGluGlnValAsnThrAspSerGluThrAla  120

Qy    433  GTTGTCAACGTCACATATGCACAGAAGAAGCAAAAATAGCCATGGACAAAATTCTACATCCCGATGAAGAGGTGAGC  492
               |||:: ||||| || |||||||||||:::||||||:::|||||||||:::|||||
Db    121  ValValAsnValThrTyrThrAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn  140

Qy    493  GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCTACATCCCGATGAAGAGGTGAGC  552
               ||| |||||||||||||||:::|||||||||||:::||||||
Db    141  GlyHisGlnLeuGlnAsnHisValLeuLysValSerTyrIleProAspGluGln-----  158

Qy    553  TCCCCTTCCCCCCCTCAGCGAGCCCAGCCGTGGGGACCACTCTTCCCGG-----GAG  603
               ||||||||||||||:::||||||||||||||||||:::|||||
Db    159  SerValGlnGlyProGluAsnGlyAsnGlyArgGlyArgGlyPheGlyAlaArgGlyAlaProArgArg  178

Qy    604  CAAGGGCCAC-----GCCCCTGGGGGCGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTG  657
               |||:: |||      :::||||||||||:::||||||||||||:::|||||
Db    179  GlnGlySerProValThrAlaGlyAlaProValLysGlnProValAspIleAspProIleProLeu  198
```

```
Db    301  LeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrLysIle  320
Qy    994  ACAATCTCATCTTTGCAGGATTTGAGCATTACAACCCCGAAAGAACCACTCGTGAAG  1053
               |||||||||||||||:::|||:::||||||||| ||||||||||||||||
Db    321  ThrIleSerProLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLys  340

Qy   1054  GGCAGCAGTTGAGGCCTGTCCCAGTGCGACATAGAGATTATGAGAAGCTGCGTGAGGCC  1113
               ||:::|||:::|||||:::||||||||||||||||||:::||||||||||:||
Db    341  GlySerIleGluProCysAlaLysAlaGluGluGlyLeuIleMetLysLysIleArgGluSer  360

Qy   1114  TTTGAAAATGATATGCTGGCTGGTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTC  1173
               :::||||||:: |||||||||:: ||||||||:::||||||||||||||||||
Db    361  TyrGluAsnAspIleAlaAlaMetAsnLeuGlnAlaHisLeuIleLeuProGlyLeuAsnLeu  380

Qy   1174  AGCCAACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACAGCAGGCCCCGC  1233
               :::|||:::|||||||||:::
Db    381  AsnAlaLeuGlyLeuPhe----------------------ProSerSerSer  391

Qy   1234  GGAGCTCCCCCGCTGCCCCCTACCACCCTTCACTACCACTCCGGATACTTCTCCAGC  1293
               :::||| :::    |||||:::----------SerThr
Db    392  GlyMetProProSerValGlyValProSerProThrSer----------SerThr  407

Qy   1294  CTGTACCCCCATCACCAGTTGGCCGTTCCCGCATCATCACTCTTATCCAGAGCAGGAG  1353
               :::||||||-------------------ProGluSerGlu
Db    408  SerTyrPro-------------------ProGluSerGlu  419

Qy   1354  ATTGTGAATCTCTCCATCCCAGGTGTGGGCGCCATCATCGGGAAGAAGAGGGGCA  1413
               |||||||||||||||:::|||||||||||||:::|||||||||||||||||
Db    420  ThrValHisLeuPheIleProAlaLeuAlaValGlyAlaIleIleGlyLysGlnGlyGln  439

Qy   1414  CACATCAAACAGCTGGCGAGATTCGCCGAGCCTCTATCAAGATTGCCCGTGCGGAAGGC  1473
               |||||:::|||||||||:::|||||||||:::||||||||||:::|||||
Db    440  HisIleLysGlnLeuSerArgPheAlaGluPheAlaSerArgIleLysSerIleAlaProAlaGluGly  459

Qy   1474  CCAGACGTCAGCGAAAGGATGGTCATCATCACCGGGCACCGGAAGCCCAGTTCAAGGCC  1533
               |||:::|||:::|||||||||||:::|||||||||||||:::|||||
Db    460  ProAspAlaLysLeuArgMetValIleIleIleThrGlyTyrProProGlnGluAlaGlnPheLysLysAla  479

Qy   1534  CAGGGACGGATCTTTTGGGAAACTGAAAGAGGAGAAACTTCTTAACCCCAAAAGAAGAGTG  1593
               ||| ||||||:::||||||||||||||||||||||:::||||||||||||
Db    480  GlnGlyArgIleTyrGlyLysLeuGlyLeuGlnPheGluGluGlyLysGlnGluVal  499

Qy   1594  AAGTGGAAGCGCCATATCAGATGCCCTCTTCCACAGTCTGCCCGGTGATTGGCAAAGGT  1653
               ||:::||||||||||||| |||:::||||||:::||||||||||||||||
Db    500  LysLeuGluThrHisIleLeuLysLysValProSerTyrAlaAlaGlyValIleGlyLysGlyGly  519

Qy   1654  GGCAAGACCGTGACGACAACTGCAGAACTAACCAGTGCAGAAGTCATCGTGCCTCGTGAC  1713
               ||||||:::|||||||||:::|||||||||||||||:::|||||||||||
Db    520  GlyLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluValValValProArgAsp  539

Qy   1714  CAAAGCCCAGATGAAATTATGGGTCAGAGAAGTGATCGTCAGAATTATCGGCTCACTCTTTTGCTAGC  1773
               |||:::|||||||:::|||||||||||||:::|||||||||:::|||||
Db    540  GlnThrProAspGluIleMetGlnArgValIleGlyValLysGlyGlyHisPheTyrAlaSer  559

Qy   1774  CAGACTGCACAGCGCCAAGATCAGGAGAATTGTACAACAGGTGAACAGCAGGAGCAGAAA  1833
               |||:::|||||||||||||:::||||||||||||||||:::|||||
Db    560  GlnLeuAlaGlnArgLysIleGlnGlyIleLeuLeuAlaGlnValArgArgGlnGlnGln  579

Qy   1834  TACCCTCAGGGAGTCGCCTCA  1854
               :::|||:::
Db    580  GlnGlnLysThrValGlnSer  586
```

```
RESULT 5
O42254
ID   O42254          PRELIMINARY;      PRT;    576 AA.
AC   O42254;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Zipcode-binding protein.
GN   ZBP1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Qy   1738  GTGATCGTCAGAATTATCGGCACTCTTCTTGCTCCAGACTGCCAGCCCAAGATCAGG  1797
                ||||||||| :::|||||||||  :::||||| |||:||||||||  |||||||
Db    534  ValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArg  553

Qy   1798  GAAATTGTACACAGGTGAAGCAGCAGCAGGAGCAGAAATACCCTGCAGGAGTCGCCTCACAG  1857
                :::|||:||  |||||||||||||||||||||||||||||||||||||||:|||
Db    554  AspIleLeuAlaGlnAlaValLysGlnGlnHisGlnLysGlyGlnSerGlyGlnLeuGlnAla  573

Qy   1858  CGGAGCAAG  1866
                ||  |||
Db    574  ArgArgLys  576
```

RESULT 6

```
ID    Q8CGX0                PRELIMINARY;        PRT;      577 AA.
AC    Q8CGX0;
DT    01-MAR-2003 (TrEMBLrel. 23, Created)
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    B-actin zipcode binding protein 1.
OS    Rattus norvegicus (Rat).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX    NCBI_TaxID=10116;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=Sprague-Dawley;
RA    Eom T., Singer R.H., Bassell G.J.;
RT    "Molecular interactions between rZBP1 and b-actin zipcode required for
RT    transport of mRNA and stimulation of spine growth.";
RL    Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF541940; AAO16210.1; -.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH_type_1.
DR    InterPro; IPR005504; RNA_rec_mot.
DR    Pfam; PF00013; KH; 4.
DR    Pfam; PF00076; rrm; 2.
DR    SMART; SM00322; KH; 4.
DR    SMART; SM00360; RRM; 2.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
DR    PROSITE; PS50102; RRM; 2.
SQ    SEQUENCE  577 AA;  63436 MW;  0647676128FBD1EE CRC64;
```

Alignment Scores:

```
Pred. No.:                                Length:          577
Score:                 6.59e-142          Matches:         395
Percent Similarity:    2023.50            Conservative:     81
Best Local Similarity: 78.94%             Mismatches:       96
Query Match:           65.51%             Indels:           31
DB:                    33.06%             Gaps:              4
                       11
```

US-09-270-437D-6 (1-3412) x Q8CGX0 (1-577)

```
Qy     73  ATGAACAAGCTTTACATCGGGAACCTGAGCCCGTCACCGGCGACGACCTCCGGCAG  132
                |||  ::   :: |||
Db      1  MetAsnLysLeuTyrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluLys   20

Qy    133  CTCTTTGGGGACAGGAGAGCTGCCCCTGGCGGACAGGGTCCTGCTGAAGTCCGGCTACCGC  192
                ::|||||   ||| |||||  :::|||    ||:||||
Db     21  ValPheAlaGluHisGlyProValIleSerIleSerTyrSerGlyLysPheGlyTyrArgAla   40

Qy    193  TTCGTGGACTACCCCAGAACTGGGCCATCCGCGCATCCGAGACCCTCCGGGTAAA  252
                |||||||| |||||||| |||||||||||
Db     41  PheValAspCysProAspSerLysTrpAlaMetLysAlaIleGluThrProSerGlyLys   60

Qy    253  GTGGAATTCGCATGGGACAAATCATGGAAGTGATTACTCAGTCTCTAAAAGCTAAGGAGC  312
                ||||||  :::||||    :::|||||||||::|||||
Db     61  ValGluLeuGlnGlyLysArgLeuGlyIleArgLeuHisSerValProLysGlnArgSer   80

Qy    313  AGGAAAAATCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTT  372
                ||||| ::: ::|||
Db     81  ArgLysIleGlnIleIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspLeu  100
```

```
Qy    658  CGGATCCTGGTCCCCACCAGTTTGTGTGTGCCCATCATCGGAAAGGAGGGCTTGACCATA  717
                |||||||||:|||||||||||||||||||||::|||||   ||::|||||
Db    199  ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGluGlyLysAlaThrIle  218

Qy    718  AAGAACATCACTACAGCAGCAGCCCAGTCCCGGTAGATATCCATAGAAAAGAGACTCTGGA  777
                :::||||||||||||:||||||||||||||:::::||||||||:::||::||||||
Db    219  ArgAsnIleThrLysGlnThrGlnSerLysIleAspIleAspArgLysArgLysAsnAlaGly  238

Qy    778  GCTGCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGACACTTCTGAAGCATGCCGC  837
                ||||:::||  :::|||:||||||||||  ||| |||
Db    239  AlaAlaAlaGluLysAlaIleLeuSerIleHisSerThrThrProGluGlyCysSerAlaAlaCysLys  258

Qy    838  ATGAATTCTTGAAATCATGCAGAAGAGGACGATGCAGACCAAACTAGCCGAAGAGATTCCT  897
                ::: |||||||||::|||||||||   |||||||::|||||||||||
Db    259  MetIleLeuGlnLeuMetGlnLysGluLysValGlyAlaLysAspThrProGlnLysLeuPro  278

Qy    898  CTGAAAATCTTGGCACACATGGCTTGGTTGGAAGACTGATTGGAAGAAAGGCAGAAAT  957
                |||||||||||   ||:::|||  ||||||||  :::::|||||   ||||:|||
Db    279  LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlyGluGlnArgAsn  298

Qy    958  TTGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCGAGGATTTG  1017
                |||||| :::||||||||::|||||| ||||||||||||||:::||||
Db    299  LeuLysLysValGluGlnIleAspThrGluThrLysThrIleThrIleSerSerLeuGlnAspLeu  318

Qy   1018  AGCATATACAACCCGAAAGAACATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGT  1077
                :::||||||||||  :::|||||||||  |||||||||  ||
Db    319  ThrLeuTyrAsnProGlnArgThrIleThrValLysGlySerIleGluAsnCysCysLys  338

Qy   1078  GCTGAGATAGAGATTATGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTCGCTGTT  1137
                |||||||   |||||||||||||  :::||||||||  |||||::::||  ||||||
Db    339  AlaGluIleGluIleMetLysLysValArgGluAlaTyrGluAsnAspValAlaAlaMet  358

Qy   1138  AACCAACAGGCAATCTGATCCCAGGGTTGAACCTCAGCGACCATTGGCATCTTTCAACA  1197
                :::  |||||||:::|||:::||| ||||||||  |||
Db    359  SerLeuGlnSerHisLeuProGlyLeuAsnLeuAlaLeuAlaValGlyValLeuPheProAla  378

Qy   1198  GGACTGTCCGTGCTATCTCCACCAGCAGGGCGCCCCGGAGCTCCCCCGCTGCCCCTAC  1257
                ||||:::|||||||||||
Db    379  SerSerAsnAlaValProPro------ProSerValSerGlyAlaAlaProTyr  396

Qy   1258  CACCCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGC  1317
                              ||||
Db    397          ------------------SerSerPheMetPro------------  401

Qy   1318  CCGTTCCGGCATCATCACTCTTATCCGGAGGATTGTGAATCTTCATCCCAACC  1377
                |||||||||:::||||||:::||
Db    402  ----------------ProGluGlnGluThrValHisValPheIleProAla  413

Qy   1378  CAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGTGGGCGAGATTC  1437
                ::::||| :::  ||||  :|:::||
Db    414  GlnAlaValGlyAlaIleAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSerArgPhe  433

Qy   1438  GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAGAGATGGTC  1497
                ||||||||:::|||||||||:::
Db    434  AlaSerAlaSerIleLysIleAlaProAlaGluGlyProGluProSerLysValArgMetVal  453

Qy   1498  ATCATCACCGGCCACCGGAGCCCAGTTCAAGGCCAGGGACGGATCTTTGGGAAACTG  1557
                :::|||||||:::||||||||  :::|||||||
Db    454  ValIleThrGlyProGluAlaGlnAlaGlnPheLysGlyArgIleThrValLysLysLeu  473

Qy   1558  AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTG  1617
                ||||||||||  ||||||||:::
Db    474  LysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluAlaHisIleArgVal  493

Qy   1618  CCCTCTTCCACAGCTGGCCGGCGGTTGATTGGACAAGTGGCAAAACGTGACGAACTGCAG  1677
                |||||| :::|||||:::||||
Db    494  ProAlaSerAlaAlaGlyArgValIleGlyLysGlyLysArgThrValAsnGluLeuGlnGln  513

Qy   1678  AACTTAACCAGTGCAGAAGTCATCGTGCCCTGTGACCAAACGCCAGAGAAAATGAGGAA  1737
                |||||:::||||:::||||||:::||||
Db    514  AsnLeuThrAlaAlaGluValValProValProGlyGlnThrProAspLysAsnAspLeu  533
```

```
Qy   373  TTGGGCTCAATATGGGACAGTTGGAGAATGTTGGAGAAGTCAACAAGCAACAGAAACCGCC  432
          ||||||||||  |||||||||||||||||||||||||||||||||||||||
Db   101  LeuAlaGlnTyrGlyThrValGlyGluAsnCysGluGlnValAsnThrGluSerGluThrAla  120

Qy   433  GTTGTCAACGTCACATATGCACAACAGAGAAGGAGCAAAAATAGCCATGAGAAGCTAAGC  492
          ||||||||||||||||:::   :::  ||||||||
Db   121  ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAsnIleAlaMetLysLeuAsn  140

Qy   493  GGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC  552
          |||||||||::|:::::     |||:||||||
Db   141  GlyHisGlnLeuGluLeuLeuSerAlaLeuLysValSerTyrIleProProAspGluGlnIleLeuAla  160

Qy   553  TCCCCTTCGCCCCCTCAGCCAGCCGTGGGACCCACTCTTCCCGGGAGCAAGGCCAC  612
          |||:::||  ::||||
Db   161  ------GlnGlyProGluAsnGlyArgArgGlyGlyPheGlySerArgGlyGlnProArg  178

Qy   613  --------------GCCCCTGGGGCCACTTCTCAGGCCAGACAGATTGATTTCCGCGTG  657
          ||:||     :::::|   :::  ||||||||
Db   179  GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnGlnAsnValAspSerIleProLeu  198

Qy   658  CGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATA  717
          |||||||||:||||||:::|||||||||||||||||
Db   199  ArgLeuLeuValProThrGlnThrGlyTyrValGlyAlaIleIleGlyLysLysGlyGlyAlaThrIle  218

Qy   718  AAGACACATCTAGGACAGACCCAGTCCCCGGGTAGAATCCATAGAAAAGAGAACTCTGGA  777
          ::|||||||||||:::||:   ||||||||
Db   219  ArgAsnIleThrLysSerGlnThrGlnSerLysIleAspSerValHisArgLysGlnValAsnAlaGly  238

Qy   778  GCTGCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGACTTCTGAAGCATGCCGC  837
          ||:||||||:::||||
Db   239  AlaAlaAlaGluLysLysAlaIleSerValHisSerThrProGluGluCysSerSerAlaCysLys  258

Qy   838  ATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCAAGAGAATTCCT  897
          ||||||||||||||||||||||||||
Db   259  MetIleLeuGluIleMetHisLysGluAlaAspGluThrLysThrAlaAspAlaLysAsnCysArg  278

Qy   898  CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGCTGATTCGAAAGAAGCAGAAAT  957
          |||||||||||:::  |||  ||||||||
Db   279  LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuLeuIleGlyLysLysGlyGluGlyIleArgGlyAsn  298

Qy   958  TTGAAGAAATTGAACATGAAACATGAGACGACCAAGATAACAATCTCATCTTTGCAGGATTTG  1017
          |:::||:::||:||||||:::   |||:||||||
Db   299  LeuLysLysValGlyGlnAspArgThrGluThrLysThrLysIleIleThrIleSerSerLeuGlnAspLeu  318

Qy  1018  AGCATATACAACCCGGAAGAAGAACCATCACTGTGAAGGGCCACAGTTGAGGCCTGTGCCAGT  1077
          :::||||||||||||||||||||||||    :: |||
Db   319  ThrLeuTyrAsnProGluArgThrIleThrValLysSerGlyAlaIleLeuGluLeuAsnCysSerArg  338

Qy  1078  GCTCAGATAGAGATTATGAAGAAGCTGCTGAGCCCTTTGAAAATGATATGTCGGCTGTT  1137
          ::||||||||||:::|| |||||||||||
Db   339  AlaGluGlnIleGluIleMetLysLysLeuLeuValArgGluLeuAlaTyrGluGlyGlyAlaValAlaAlaMet  358

Qy  1138  AACCAACAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTCTTTTCAACA  1197
          :::   ||:|||||||||||||||||||
Db   359  SerLeuGlnSerHisSerLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla  378

Qy  1198  GGACTGTCCGTGCTATCCTCACCACAGGGCCCGCGGAGCTGCCCCCCTAC  1257
          |||::: ::|||||  |||
Db   379  SerSerSerAlaValProProPro------ProSerSerValThrGlyValAlaAlaProTyr  396

Qy  1258  CACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGC  1317
          ||||||||:::  |||
Db   397  GlySerPheMet------------------------------------------------  400

Qy  1318  CCGTTCCCGGCCATCATCACTCCTTATCCAGAGCAGGAGATTGTGAATCCTCCAACC  1377
          ::|  :::|||||||||:::||
Db   401  --------------------GlnAlaProGlyGlyLeuMetValGlnIleValPheIleProAla  414

Qy  1378  CAGGCTGTGGGGGCCCCATCATCGGGAAGAGGGGGCCACACATCAAAACAGTCTGGCGAGATTC  1437
          |||||||||||||||||||||||||||:::|||
Db   415  GlnAlaValAlaValGlyAlaIleIleGlyLysGlyGlyGlnHisIleLysGlnLeuSerArgPhe  434
```

```
Qy  1438  GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAGGGCCCAGAGTCAGCGAAGGATGGTC  1497
          ||  ||||||||||||||||||||||||||||  |||||||||
Db   435  AlaSerAlaSerIleLysIleAlaProGluThrProAspSerLysValArgMetVal  454

Qy  1498  ATCATCACCGGGCCACCGGAGCCCAGTTCAAGGCCCAGTTCAAGGCCGGACGGATCTTTGGGAAACTG  1557
          :::|||||||||||||||||:::|||||||||||||||
Db   455  ValIleThrGlyProProGluAlaGlnPheLysAlaGlnPheLysAlaGlyArgIleTyrGlyLysLeu  474

Qy  1558  AAAGAGGAGAAACTTCTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGCGCATATCAGAGTG  1617
          ||||||||||||||||:::||||||||||||||||||
Db   475  LysGluGluLysLeuLeuAsnProLysGluGluValLysLeuGluThrHisIleArgVal  494

Qy  1618  CCCTCTTCCACAGCTGCGCCGGGTGATTGGCCAAGGTGGCAAGACCGTGAACGAACTGCAG  1677
          |||:||||  ::|||||||||||||||||||||||
Db   495  ProAlaSerAlaLeuAlaGlyLysArgValIleIleGlyGlyGlyLysLysThrValAsnGluLeuGln  514

Qy  1678  AACTTAACCAGTGCAGAAGTCATCGTGCCTCCGACAGCCCAGATGAAAATAGAGAA  1737
          |||||||||||||||||||||:::  |||  ||||||
Db   515  AsnLeuThrAlaAlaGluValValProThrAlaGlnMetProArgAspGluAsnAspGln  534

Qy  1738  GTGATCGTCAGAATTATCGGGACCACTCTTTTGCTAGCCGACGATCGCAGCGGCAAGATCAGG  1797
          ||||||||||:::||||||||||||||
Db   535  ValIleValLysIleLeuGlyGluHisGlyTyrHisPheTyrAlaSerGlnMetAlaGlnAsnArgLysIleArg  554

Qy  1798  GAAAATTGTACAACAGTGAAGCAGCGGAAGCGGCAGCAGAATACCCTCAGGGAGTGCGCTCCACAG  1857
          :::||||||||||:::||||||||||
Db   555  AspIleLeuGluAlaGlnIleValLysGlyGlyGlnIleHisGlnGlnLysLysGlyGlnSerAsnGlnAlaGly  574

Qy  1858  CGCAGCAAG  1866
          ||||||||||
Db   575  ArgArgLys  577
```

RESULT 7
O88477
ID   O88477   PRELIMINARY;   PRT;   577 AA.
AC   O88477;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Coding region determinant binding protein.
GN   IGF2BP1 OR CRDBP.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9217743; PubMed=1559612;
RA   Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT   "Control of c-myc mRNA half-life in vitro by a protein capable of
RT   binding to a coding region stability determinant.";
RL   Genes Dev. 6:642-654(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94158886; PubMed=8114742;
RA   Herrick D.J., Ross J.;
RT   "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RT   influence of the coding and 3' untranslated regions and role of
RT   ribosome translocation.";
RL   Mol. Cell. Biol. 14:2119-2128(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9179348; PubMed=8132663;
RA   Prokipcak R.D., Herrick D.J., Ross J.;
RT   "Purification and properties of a protein that binds to the C-terminal
RT   coding region of human c-myc mRNA.";
RL   J. Biol. Chem. 269:9261-9269(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322234; PubMed=9178888;
RA   Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RA   Gruppuso P.A., Ross J.;
RT   "Developmental regulation of CRD-BP, an RNA-binding protein that

```
RT    stabilizes c-myc mRNA in vitro.";
RL    Oncogene 14:1279-1286(1997).
RN    [5]
RP    SEQUENCE FROM N.A.
RA    Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RA    Fleisig A.J.;
RL    Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN    [6]
RP    SEQUENCE FROM N.A.
RC    STRAIN=C57BL/6J; TISSUE=Embryo;
RX    MEDLINE=21085660; PubMed=11217851;
RA    Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA    Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA    Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA    Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA    Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA    Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA    Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA    Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA    Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA    Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA    Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA    Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA    Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA    Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA    Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA    Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA    Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA    Hayashizaki Y.;
RT    "Functional annotation of a full-length mouse cDNA collection.";
RL    Nature 409:685-690(2001).
CC    -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR    EMBL; AF061569; AAC72743.1; -.
DR    EMBL; AK013940; BAB29071.1; -.
DR    HSSP; P11940; 1CVJ.
DR    MGD; MGI:1890057; Igf2bp1.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH type 1.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfan; PF00013; KH; 4.
DR    Pfan; PF00076; rrm; 2.
DR    SMART; SM00322; KH; 4.
DR    SMART; SM00360; RRM; 2.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
DR    PROSITE; PS50102; RRM; 2.
SQ    SEQUENCE  577 AA;  63450 MW;  EFB1AF2FF9F0344 CRC64;
```

```
Alignment Scores:
Pred. No.:                              Length:       577
Score:             9.29e-142            Matches:      394
Percent Similarity:      78.94%         Conservative: 82
Best Local Similarity:   65.34%         Mismatches:   96
Query Match:             33.03%         Indels:       31
DB:                                     Gaps:         4

US-09-270-437D-6 (1-3412) x O88477 (1-577)
```

```
Qy     73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGTCACCGCGCGACACCTCCGGCAG  132
          ||||||||||||| ||||||||||  ||:::||||||||||||||  :::
Db      1 MetAsnLysLeuTyrIleGlyAsnLeuSerValThrProAlaSerLeuAspLeuProAla   20

Qy    133 CTCTTTGGGGACAGGAAGTGCCCTGGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
          :::||||||||| ||||||  :::||||||||||||||||||||:::||
Db     21 ValPheAlaGluHisLysIleSerTyrSerGlyGlnPheLeuValLysSerGlyTyrAla   40

Qy    193 TTCGTGGACTACCCCCACCAGAACTGGGCCATCCCGCCCATCGAGACCCTCCGGGTAAA  252
          ||||||||||||||||||  ::||||||||||||||||||||||||||||||
Db     41 PheValAspCysProAspGluHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys   60

Qy    253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC  312
          ||||||  ||||:::||| |||||||||  ||||| ||||:::||||
Db     61 ValGluLeuGlnGlyLysArgLeuGluIleAspTyrSerValProLysLysValProArgSer   80
```

```
Qy    313 AGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCCAGTGGAGGTGTTGGATGGACTT  372
          ||||||||||||||||  :::|||||||||| |||||||
Db     81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspSerLeu   100

Qy    373 TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAACCCCC  432
          |||||||||||||||||||||||||:::|||||||||||||
Db    101 LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValAsnThrGluSerGluThrAla   120

Qy    433 GTTGTCAACGTCACATATGCAACAAGAGAAGAGAGCAAAAATAGCCATGGAGAAGCTAAGC  492
          |||||||||||||:::||  ::::       |||||||||||||
Db    121 ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn   140

Qy    493 GGGCATCAGTTTGAGACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC  552
          ||||||||||||||||  :::||||||||||||||||||||||||
Db    141 GlyHisGlnLeuGluAsnHisAlaLeuLysValSerTyrIleProIleProAspGlnIleThr   160

Qy    553 TCCCCTTGCCCCCCTCAGCGAGCCCAGCTGGGGACCACTCCTTCCCGGGACGCAGGCCAC  612
          ||:::||||||||||||||  :::|||
Db    161 ------GlnGlyProGlnLysArgArgGlyGlyPheGlySerArgGlyPheGlnProArg   178

Qy    613 --------------------------GCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCGGCTG  657
          ------                    |||||  ::: ::::|||||||
Db    179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValAspProIleProLeu   198

Qy    658 CGGATCCTGGTCCCCACCCAGTTTGGTGCGCCATCATGGCATCATGGCAGAAGGAGGGCTTGACCATA  717
          ||||||||||||||||||||  :::|||||||||||||||
Db    199 ArgLeuLeuValProThrArgGlnTyrValGlyValAlaIleIleGlyAlaThrIle   218

Qy    718 AAGAACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGA  777
          :::||||||||||||  :::|||||||||||||||||||:::||
Db    219 ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysValGluAsnAlaGly   238

Qy    778 GCTGCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGACTTCTGAAGCATGCGCGC  837
          ||||||||||||||||  :::||||||||||||||
Db    239 AlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys   258

Qy    838 ATGATTCTTGAAATCATGCAGAAAGAGCCAGATGAGACAAACTAGCCGAAGAGATTCCT  897
          ||||||||||||||||||||  :::|||||||||||||||
Db    259 MetIleLeuGluIleMetHisLysGluAlaLysAspThrLysThrArgAlaAsnGluValPro   278

Qy    898 CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAAGGCAGAAAT  957
          |||||||||||  :::|||||||||||||||||
Db    279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlyGlyArgAsn   298

Qy    958 TTGAAGAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTG  1017
          ||||||||||||||||  :::|||||||||||||
Db    299 LeuLysLysValGluGlnLeuAspThrGluThrLysIleThrIleSerSerLeuGlnAspLeu   318

Qy   1018 AGCATATACAACCCGGAAAGAACCATCACTGAGGGCACAGTGAGCCTGTGCCAGT  1077
          ||:::||||||||||||||||||||  :::
Db    319 ThrLeuTyrAsnProGluArgThrIleThrGluGlyThrValSerLeuCysSerArg   338

Qy   1078 GCTCAGATAGAGATTATGAAGAGCTGCCTGAGGCCTTTGAAAATGATGATGCTGGCTGGTTGTT  1137
          |||||||||||||||||||  :::-----||||
Db    339 AlaGlnIleGluIleMetLysSerValArgGluAlaTyrGluValGluAlaAlaAlaMet   358

Qy   1138 AACCAACAAGCCAATCTGATCCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACA  1197
          ||:::                ||||||||||||  :::
Db    359 SerLeuGlnSerHisLeuProGlyLeuAsnLeuAlaLysAlaValGlyLysLeuPheProAla   378

Qy   1198 GGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCCCGGAGCTCCCCGTGCCCCTAC  1257
          ||||||||||  ::: ||-----
Db    379 SerSerSerAlaValProPro------ProSerSerValThrGlyValAlaIleProTyrAla   396

Qy   1258 CACCCCTTCACTACCCACCTCCGGATACTTCTCCAGCGTGTACCCCATCACCAGTTTGGC  1317
          ||||||||||||||||||||||||||||
Db    397 SerSerPheMet-----------------------------------   400

Qy   1318 CCGTTCCCGCCATCATCACTTTATCCAGCAGGAGAGATTGTGAATCTTCATCCCAACC  1377
          ||||||  ::: ||||||||
Db    401 ------------------------GlnAlaProGluGlnGluMetValGlnMetPheLeuValProAla   414
```

```
Pred. No.:
Score:                        2.6e-141
Percent Similarity:           2015.50
Best Local Similarity:        78.61%
Query Match:                  65.34%
Db:                           32.93%
                              4

Length:        577
Matches:       394
Conservative:  80
Mismatches:    98
Indels:        31
Gaps:          4
```

```
US-09-270-437D-6  (1-3412) x Q9NZI8  (1-577)
```

```
Qy    73  ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCCGCCGTCACCGCCGACGACTTCCGGCAG  132
              :|||||||||||||||||||||||||||||||||||  ::||::
Db     1  MetAsnLysLeuTyrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluGluLys  20

Qy   133  CTCTTTGGGGACGAGGAGCTGCCCCTGGGCGGACAGCAGGTGCTCGCTGAAGTCCGGCTACGCC  192
              :||| |||:     ::||||    ::||||||
Db    21  ValPheAlaGluHisLysIleSerTyrSerGlyGlnPheLeuValLysSerGlyTyrAla  40

Qy   193  TTCGTGGACTACCCCGACCAGACTGGGCCATCGGGCCCCATCGAGACCCTCTCGGGTAAA  252
              |||:||| ||||||||  ||||||||||||
Db    41  PheValAspCysProAspGluHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys  60

Qy   253  GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC  312
               ||::||||||||  :||||||||||||||||||||||||
Db    61  ValGluLeuGlnGlyLysArgLeuGluIleGluHisSerValProLysLysGlnArgSer  80

Qy   313  AGGAAAATTCAGATTGCAAACATCCCTCTCCACCTGCAGTGGGAGGTGTTGATGCAGCTT  372
              ||||||||||||||  |||||||:|||||||||
Db    81  ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuLeuAspSerLeu  100

Qy   373  TTGGCTCAATATGGACAGTGGACAGATGTGGAACAGTCAACACAGAACCGCC  432
              :::||||      |||||
Db   101  LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValAsnThrGlueSerGluThrAla  120

Qy   433  GTTGTCAACGTCACATATGCAACAAGACGAAGAGCAAAAATAGCCATGGAAGAGTAAGC  492
              ||||||||        :||   :::||||||
Db   121  ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn  140

Qy   493  GGGCATCCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGATGAAGAGGTGAGC  552
              |||||||||||    ::|:::: |:||||||
Db   141  GlyHisGlnLeuGluAsnHisAlaLeuValSerTyrIleProGlnIleAla  160

Qy   553  TCCCCTTGCCCCCCAGCGAGCCAGCTCGTGGGGACCACTCTTCCCGGGAGCAAGGCCAC  612
              |||     :::||
Db   161  ------GlnGlyProGluAsnGlyArgArgGlyGlyTyrPheGluGlyGlySerArgGlyGlnProArg  178

Qy   613  --------------GCCCCTGGGGACCACTTCTCAGGCCAGACAGATTGATTTCCCGCTG  657
                             :||| ||| :::::||||
Db   179  GlnGlySerProValAlaAlaGlyAlaProAlaAlaLysGlnGlnInValAspIleProLeu  198

Qy   658  CGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATA  717
              |||||| :::::||||:::||||||
Db   199  ArgLeuLeuValProThrGlnTyrValGlyValAlaIleIleGluGlyLysAlaThrIle  218

Qy   718  AAGAACATCACTAAGCAGACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACTCTGGA  777
              :::||||||||||||||||||  ||:||:||
Db   219  ArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGly  238

Qy   778  GCTGCAGAGAGCCTGTCACCATCCATGCCACCCCAGGGGGACTTCTGAAGCATGCCGC  837
              ||| :::::|||||||::::  :::||||
Db   239  AlaAlaGluGluLysAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys  258

Qy   838  ATGATTCTTGAAATCATGCAAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAATTCCT  897
              ||||||:::::|||||||||||
Db   259  MetIleLeuGluIleMetHisLysGluAlaGluAlaLysAspThrLysThrAlaGluValPro  278

Qy   898  CTGAAAATCTTGGCACCAATGGCTTGGTTGGAAGACTGTTGGAAAAGAAGAGGCAGAAAT  957
              |||||:::||:::
Db   279  LeuLysThrLeuAlaHisAsnAsnPheValGlyArgLeuLleGlyLysArgAsn  298

Qy   958  TTGAAGAAATTGAACATGAACAAGATAACAATCATCTTTGCAGGATTTG  1017
              |||||:|  :::||||
Db   299  LeuLysLysValGlyValGlnAspThrGlyThrLysIleThrIleSerLeuLysGlnAspLeu  318
```

```
Qy  1378  CAGGCTGTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTC  1437
              :|||||||||||||||||||||||||||||||||||:::|||||
Db   415  GlnAlaValGlyValAlaIleIleGlyLysLysGlyHisIleLysGlnLeuLysGlnLeuSerArgPhe  434

Qy  1438  GCCGAGCCCTCTATCAAGATTGCCCCTGCCGGACGTGCAGCGCCCGAAGCGAAAGGATGGTC  1497
              ||||||||||||||||||||||||||||||||||||||||:||
Db   435  AlaSerAlaSerIleLysIleAlaProAlaProGluThrProAspSerLysValArgMetVal  454

Qy  1498  ATCATCACCGGCGCACCGGAGCCCAGTTCAAGGCCCAGGGACGGACGGATCTTTGGGAAACTG  1557
              :::||||||||||||||||||||||||:::||||||||||
Db   455  IleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyLyArgIleLeuTyrGlyLysLeu  474

Qy  1558  AAAGAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCCGCATATCAGAGTG  1617
              ||||||||||||||||||||||||||||||||||||||||
Db   475  LysGluGluAsnPhePheGlyProLysGluValLysLeuGluThrHisIleArgGlyVal  494

Qy  1618  CCCTCTTCCACAGGTGCCCGGGTGATTGGCAAAGGTGGCAAAGACCGTGAAGCGTGCAG  1677
              |||:::||||||||||||||||||||||||||||||||||||
Db   495  ProAlaSerAlaAlaGlyAlaGlyValIleGlyLysLysGlyGluGlnIleAsnGluLeuGln  514

Qy  1678  AACTTAACCAGTGCAGAAGTCATCGTGCCCTCGTGACCAAAACGCCAGATGAAAATGAGGAA  1737
              |||::::||||||||:||:||||||||||:::::||
Db   515  AsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsnAspGln  534

Qy  1738  GTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACGCGCCAAGATCCAGG  1797
              |||||||||||||||||||||||||||||||||
Db   535  ValIleValLysIleIleGlyHisPheTyrAlaSerGlnInMetAlaGlnArgLysIleArg  554

Qy  1798  GAAATTGTACAACAGGTGAAGCAGCAGAGCAGAAAATACCCTCAGGGAGTCGCCTCACAG  1857
              :::::||||||||||||
Db   555  AsnIleLeuAlaGlnInValLysGlnInHisGlnLysGlyGlnInSerAsnAlaAlaGlnAla  574

Qy  1858  CGCAGCAAG  1866
              ||||||
Db   575  ArgArgLys  577
```

```
RESULT 8
Q9NZI8
ID   Q9NZI8         PRELIMINARY;      PRT;    577 AA.
AC   Q9NZI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   mRNA-binding protein CRDBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA   Panoutsakopoulos G., Kyriazoglou I., Voutoulias S., Tsiapalis C.M.,
RA   Kittas C., Agnantis N., Pandis N.;
RT   "Ectopic expression of a KH-domain containing protein, highly
RT   homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT   malignant mesenchymal tumors.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF198254; AAF37203.1; -.
DR   HSSP; P11940; 1CVJ.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE  577 AA;  63456 MW;  0749A060F252D81D CRC64;

Alignment Scores:
```

```
[1]
RN
RP    SEQUENCE FROM N.A.
RC    STRAIN=C57BL/6J;
RX    MEDLINE=22354683; PubMed=12466851;
RA    The FANTOM Consortium,
RA    the RIKEN Genome Exploration Research Group Phase I & II Team;
RT    "Analysis of the mouse transcriptome based on functional annotation of
RT    60,770 full-length cDNAs.";
RL    Nature 420:563-573 (2002).
DR    EMBL; AK044850; BAC32119.1; -.
DR    MGD; MGI:1890357; Igf2bp1.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR004087; KH_dom.
DR    InterPro; IPR004088; KH_type_1.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfam; PF00013; KH; 4.
DR    Pfam; PF00076; rrm; 2.
DR    SMART; SM00322; KH; 4.
DR    SMART; SM00360; RRM; 2.
DR    PROSITE; PS50084; KH_TYPE_1; 4.
DR    PROSITE; PS50102; RRM; 2.
SQ    SEQUENCE 577 AA; 63378 MW; D75897450841532E CRC64;

Alignment Scores:
Pred. No.:              3.09e-141        Length:       577
Score:                  2014.50          Matches:      393
Percent Similarity:     78.77%           Conservative: 82
Best Local Similarity:  65.17%           Mismatches:   97
Query Match:            32.91%           Indels:       31
DB:                     11               Gaps:         4

US-09-270-437D-6 (1-3412) x Q8BRH1 (1-577)

Qy    73  ATGAACAAGCTTTACATGGGACCTGAGCCCCGGCCGTCACCCGCCGAGACCTTCGGCCAG 132
                                                               :::||| |||
Db     1  MetAsnLysLeuTyrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluLys 20

Qy   133  CTCTTGGGGACAGGAGAAGCTGCCCTGCGCGGGGACAGGTCCTGCTGAAGTCCGGCTACGCC 192
           ::||||        :||:|| |||||
Db    21  ValPheAlaGluHisLysIleSerTyrSerGlyGlnPheLeuValLysSerGlyTyrAla 40

Qy   193  TTCCTGGACTACCCCGACCAGAACTGGGCCATCGGCCATCGAGACCCTCCGGGAAA 252
           ||||||||     |||||||     :||:|| |||
Db    41  PheValAspCysProAspGluHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

Qy   253  GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTTAAGGAGC 312
           ||||||||    :||||||||     ||||
Db    61  ValGluLeuGlnGlyLysArgLeuGluLeuHisSerValProLysGlnThrGlnAsnGln 80

Qy   313  AGGAAAATTCAGATTCGAAACATCCTCCTCACCTGCAGTGGGAGGTGTGGATGGGACTT 372
           |||||||| ||||||||     :|||:::
Db    81  ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspSerLeu 100

Qy   373  TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAAAACCCGCC 432
           ::|||        |||||||     ||||
Db   101  LeuAlaGlnTyrGlyThrValGluAsnCysGluGlnValAsnThrGluSerGluThrAla 120

Qy   433  GTTGTCAACGTCACCATATGCAACAAGAGAAGAGAAGCAAAATAGCCATGGAGAAGCTAAGC 492
           ||||||        ::   :::   |||
Db   121  ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn 140

Qy   493  GGGCATCAGTTGAGAACTACTCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC 552
           |||||||||        |||:::|||||   ::|:||||
Db   141  GlyHisGlnLeuGluAsnHisAlaAlaLeuLysValSerTyrIleProAspGluIleThr 160

Qy   553  TCCCCTTGCCCCCCTCCAGCGAGGCTGGGGACCACTCTTCCCGGGAGCAAGGCCAC 612
           ||||| ::|                |||||||
Db   161  ------GlnGlyProGluAsnGlyArgArgGlyGlyPheGlySerArgGlyGlnProArg 178

Qy   613  -----------------------------------GCCCCTGGGGGCACTTCTCAGGCTCCAGCACAGATTGATTTCCCGCTG 657
                                               ||| ::: :::|||
Db   179  GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnProValAspAlaIleProLeu 198
```

```
Qy  1018  AGCATATACAACCCGGAAAGAAACCATCGTGAAGGGCACAGTTGAGGCCGTGCCAGT 1077
           :::::||||||||||||||||||||||||||||||||    ::||| ||
Db   319  ThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIleGluAsnCysCysArg 338

Qy  1078  GCTGAGATAGAGATATGAAGAAGCTGCGTGGAAATGATGCTGGCTGTT 1137
           ||||||||||||||||||||||||||||||||||   :::|||
Db   339  AlaGluGlnGluIleMetLysLysValArgGluAlaTyrGluAsnAspValAlaAlaMet 358

Qy  1138  AACCAACAAGCCAATCTGATCCCCAGGGTTGAACCTCAGGGCACTTGGCATCTTTTCAACA 1197
           :::                ||||||::||||||
Db   359  SerLeuGlnSerHisLeuThrProGlyLeuAsnLeuAlaValAlaValGlyLeuProheProAla 378

Qy  1198  GGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCCCGGAGCTCCCCCGCTGCCCCCTAC 1257
           ||| ::   ||||| :::|||       ||||||
Db   379  SerSerSerAlaValProPro------ProSerSerValThrGlyAlaGlyAlaProTyr 396

Qy  1258  CACCCCTTCACTACCCCACTCCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGC 1317
           |||||||||||||||||    :||||||||||
Db   397  SerSerPheMet------------------------------------------------- 400

Qy  1318  CCGTTCCCGGCCATCATCATCTTTATCCAGCAGCAGGAGATTGTGAATCTTCTTCATCCCAACC 1377
                                ||||||||||    ::|||||||||
Db   401  -----------------------GlnAlaProGluGlnGluMetGluValGlnValPheIlePheIleProAla 414

Qy  1378  CAGGCTGTGGCCCATCATCGGGAAGAAGAGGGGGCCACACATCAAACAGCTGGGCGAGATTC 1437
           ||| :::        ||||||||| ||||    :::|||
Db   415  GlnAlaValGlyTyrAlaIleIleLeuGlyLysLysGlyGlnHisIleLysGluserArgPhe 434

Qy  1438  GCCGGAGCCTCTATCAAGATGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGAAAGATGGTC 1497
           ||||| |||| ||||||||||||| |||||||
Db   435  AlaSerAlaSerIleLysIleLeuAlaProProGluProGluThrProAspSerLysValArgMetVal 454

Qy  1498  ATCATCACCGGGCCACGGAAGCCCAGTTCAAGGCCAGGGACGGATCTTTGGGAAACTG 1557
           ||||||||||||||||||||||||    :::||||||
Db   455  IleIleThrGlyProGluProProGluAlaGlnPheLysArgIleGlyIleTyrGlyLysLysLeu 474

Qy  1558  AAAAGGAAAACTCTTTACCCCAAAGAAGAAGAAGTGAAGCTGAAGCGCATATCAGAGTG 1617
           ||||||||||||||||    ||||||||| ||||
Db   475  LysGluGluLeuAsnPheGlyProLysGluProLysAspAlaLysValLeuGluThrHisSerIleArgVal 494

Qy  1618  CCCCTTCCACAGCTGGCCGGCTGATTGGCCGGGTGATTGGCAAGAGTGGCAAGACGACTGCAG 1677
           ||::|||               |||| |||||||
Db   495  ProAlaSerAlaAlaGlyValIleGlyValLysGlyLysThrValAsnGluLeuGln 514

Qy  1678  AACTTAACCAGGTGAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAA 1737
           |||||||||||||||               :::|||
Db   515  AsnLeuThrAlaAlaGluValProLysValProArgAspArgGlnThrProGluAsnAspGln 534

Qy  1738  GTGATCGTCAGAATTATCGGGCACTTCTTTGCTACGCAGACTGCACAGGCCAGATCAGG 1797
           |||||||| ||||||||               :||||
Db   535  ValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLysIleArg 554

Qy  1798  GAAAATTGTACAACAGTGAAGCAGCAGGAGAGAAATACCCTCAGGGAGTCGCCTCACAG 1857
           :::|||||        |||||||
Db   555  AspIleLeuAlaGlnAsnLysLysGlnIleHisGlnLysSerGlyGlyLysGlnGlnAsnGlnAlaGlnAla 574

Qy  1858  CGCCAGCAAG 1866
           ||| |||
Db   575  ArgArgLys 577

RESULT 9
Q8BRH1
ID   Q8BRH1     PRELIMINARY;     PRT;    577 AA.
AC   Q8BRH1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Insulin-like growth factor 2.
GN   IGF2BP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
Db   535 ValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlyAlaArgLysIleArg 554
         :::||| |||||||||::  |||||||||||||| ::||
Qy  1798 GAAATTGTACACACAGTGAAGCAGTGAAGCAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAG 1857
         :::||:|| |||||  |||||||||||||   :::
Db   555 AspIleLeuAlaGlnValLysGlnHisGlyGlnSerAsnLeuAlaGlnInAla 574

Qy  1858 CGCAGCAAG 1866
         |||
Db   575 ArgArgLys 577
```

RESULT 10
Q80US9

```
ID   Q80US9         PRELIMINARY;      PRT;   577 AA.
AC   Q80US9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Igf2bp1 protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC051679; AAH51679.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 2.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE 577 AA; 63378 MW; D439F544257DA3CE CRC64;
```

Alignment Scores:

Pred. No.:		Length: 577
Score:	3.09e-141	Matches: 393
Percent Similarity:	78.77%	Conservative: 82
Best Local Similarity:	65.17%	Mismatches: 97
Query Match:	32.91%	Indels: 31
DB:		Gaps: 4

```
Qy   658 CGGATCCTCGGTCCCCACCCAGTTCTTGGTGCCATCATCGGAAAGAGGGGGCTTGACCATA 717
         ||| ::::|||||::|| |||||||::||||
Db   199 ArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLysGluGlyLeuAlaThrIle 218

Qy   718 AAGAACATCACTAAGCAGCACCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGA 777
         |||||||||||||||||||||
Db   219 ArgAsnIleThrLysGlnHisSerLysAspValHisAspLysHisArgLysAsnAlaGly 238

Qy   778 GCTCGAGAGAAGCCTGTCACCATCATGCCGAGGGGACTTCTGAGCATGCCGC 837
         ||||||||||| :::::|||||:: ||| ||| |||||:::
Db   239 AlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCysSerSerAlaCysLys 258

Qy   838 ATGATTCTTGAAATCATGCAGAAAGAGAGGGCAGATGAGACCAAACTAGCCGAAGAGATTCCT 897
         ||||||||||||| :::||  |||||||||||||| ::::|||:::||
Db   259 MetIleLeuGluIleMetHisLysLysGluArgLysThrLysAspThrAlaAspHisGluValPro 278

Qy   898 CTGAAAAATCTGGCACACAATGGCTTGGTTGGTGAAGACTGATTGGAAAAGAAGGCAGAAAT 957
         |||||||||||||||::|||:: |||||||||||||||||
Db   279 LeuLysIleLeuLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsn 298

Qy   958 TTGAGAAAATTGAACATGAAACAGGGGACCAAGATAACAATCCATCTTTGCAGGATTTG 1017
         ||||:::|| :::||:: |||||||:: ||| ||||||::
Db   299 LeuLysLysValGlyGluGlnThrGluThrTyrIleSerIleThrIleSerLeuGlnAspLeu 318

Qy  1018 AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGT 1077
         ::::||||:: |||::|||||||||||||:: |||||||||||||::||
Db   319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaValGluAlaCysCysArg 338

Qy  1078 GCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAGAAAATGATATGCTGGCTGTT 1137
         ||||||||||||||||| ||||||||||||||
Db   339 AlaGluGluGlnIleGluIleMetLysLysLeuArgGluAlaTyrGluAsnAspValAlaAlaMet 358

Qy  1138 AACCAACAAGCCAATCTGATCCCCAGGGGTGAACCTCAGCCGACACTTGGCATCTTTTCAACA 1197
         ::: ||||:::||| ::::|||||||||||::
Db   359 SerLeuGlnSerHisLeuIleProArgGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla 378

Qy  1198 GGACTGTCCGTGCTATCTCCGTCACCACCAGGGCCCGGGAGCTCCCCGGGCCCCCCTAC 1257
         :::  |||||||:: |||     |||||||:::
Db   379 SerSerSerAlaValProProPro-----ProSerSerValThrGlyAlaAlaProTyr 396

Qy  1258 CACCCCTTCACTACCCCACTCCGGATACTTCTCCGAGCCTGTACCCCCATCACCAGTTTGGC 1317
         |||
Db   397 SerSerPheMet------------------------------------------------- 400

Qy  1318 CCGTTCCCCCATCATCACTCCTTATTCCCAGACCACGAGATTGTGAATCTCTTTCATCCCCAACC 1377
         |||::           ::::|||||||||||||
Db   401 -------------------GlnAlaProGluGlnInGlyMetValGlnInValPheIleProAla 414

Qy  1378 CAGGCTGTGGCGCCCATCATCGGGAGAAGAGGGGGCCACACATCAAACAGTCGCGAGATTC 1437
         |||||||||||||||||||||||
Db   415 GlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnInLeuSerArgPhe 434

Qy  1438 GCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCAGAGCGTCAGCGAAAGGATGGTC 1497
         ||::||||||||||||||||
Db   435 AlaSerAlaSerIleLysIleAlaProProGluThrProAspSerArgLysValArgMetVal 454

Qy  1498 ATCATCACCGGGCCACCGGGAAGCCCAGTTCAAGCGGACGGATCTTTGGGAAACTG 1557
         ::||||||||||||||||||||||
Db   455 ValIleThrGlyProGluLysGlyProGlnInAlaGlnInPheLysAlaGlnInGlyArgIleTyrGlyLysLysLeu 474

Qy  1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGAAGCGCATATCAGAGTG 1617
         |||||||||||||||||||
Db   475 LysGluGluGluAsnPhePheGlyProLysGluGluSerGluAlaGluAlaHisIleArgVal 494

Qy  1618 CCCTTCTTCCACAGCTGGCCGGGTGATTGGCAAAGGTGCAAGACCGTGAACGAACTGCAG 1677
         ||||||||||||
Db   495 ProLeuSerAlaAlaGlyAlaArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln 514

Qy  1678 AACTTAACCAGTGCAGAAGTCACCGTGCCTGTGACCAAACCGCCAGATGAAAATGAGGAA 1737
         ||::::|||||:: |||||||
Db   515 AsnLeuThrAlaAlaGluValThrValProValThrLysProProAspGluAsnAspGlnGln 534

Qy  1738 GTGGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGATCGCACAGACTGCACAAGATCAGG 1797
```

US-09-270-437D-6 (1-3412) x Q80US9 (1-577)

```
Qy    73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGCCGACCTTCCGGCAG 132
         ||||||||||||||||||||||||:::      :::||||||      :::
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuAsnGluSerValThrProAlaAspLeuGluLys 20

Qy   133 CTCTTTGGGGACAGGAAGTGCCCCTGGCCGGGACAGGTCCTGCTGAAGTCCGGCTACGCC 192
         ||||||   ||||::   |||||:   ||||||||     ::|||:||||
Db    21 ValPheAlaGluHisLysIleSerTyrSerGlyGlnPheLeuValLysSerGlyTyrAla 40

Qy   193 TTCGTGGACTACCCGACCAGGAACTGGGCGCCATCGGCGCCATCGAGACCCTCGGGTAAA 252
         |||||      |||||||:|:||                ||||||  :||||||
Db    41 PheValAspCysProAspGluHisTrpAlaMetLysAlaIleGluThrPheSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGAAGTTGATTACTCAGTCTCTAAAAAGTCTAAGGAGC 312
         ||||||  ||||||      :::|||||||:|||
Db    61 ValGluLeuGlnGlyLysIleMetLysLeuIleSerValProLysLysLysGlnArgSer 80

Qy   313 AGGAAAATTCAGATTCGAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
         ||||::||||   |||||||    :::|||||
Db    81 ArgLysIleGlnIleArgAsnIleProProGlnLeuArgTrpGluValLeuAspGlyLeu 100

Qy   373 TTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACAGACACAGAAACCGGC 432
         ||||||||||:::|||||||::|||||||::|::||||
Db   101 LeuAlaGlnTyrGlyThrValGluAsnValAsnLysSerGluThrAla 120

Qy   433 GTTGTCAACGTACATATGCAACAAGAAGAAGCAAAATAGCCATGAGAAGCTAAGC 492
         ::||||||||||:::||||||||||:||||
Db   121 ValValAsnValThrTyrSerAsnArgGluGlnThrArgGlnAlaIleMetLysLeuAsn 140

Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGATGAAGAGGTGAGC 552
         |||||||||:   |||:::|||||
Db   141 GlyHisGlnLeuGluAsnHisAlaLeuLysValSerTyrIleProGluProAspGlnIleThr 160

Qy   553 TCCCCTTCGCCCCCTCAGCGTGGGAGCCCAGTCTTCCCGGAGCAAGGCCAC 612
         ||||:::      :::|||||||
Db   161 ------GlnGlyProGluAsnGlyLyArgArgGlyGlyPheGlySerArgGlyGlnProArg 178

Qy   613 -------------------------GCCCCTGGGGCACTTCTCAGGCCCAGACAGATTGATTTCCCGCTG 657
                                   |||||||      :::      |||::||
Db   179 GlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGlnProValAspProIleProLeu 198

Qy   658 CGGATCCTGGTGCCCCACCCAGTTTCTTGGTTGCCATCATGGAAGGAGGGCCTTGACCATA 717
         |||:::||   ||||||||||||||   :::||||
Db   199 ArgLeuLeuValProHisProIleTyrValGlyAlaIleIleGluGlyLysGluGlyAlaThrIle 218

Qy   718 AAGGAACATCACTAAGCAGCAGACCCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGA 777
         ::||||||||:::      ||||||||||      :::||||
Db   219 ArgAsnIleThrLysGlnThrGlnSerLysIleSerAspValHisArgLysGluAsnAlaGly 238

Qy   778 GCTGCAGAGAAGCCCTGCACCATCCATGCACCCCAGAGGGACTTCTGAAGCATGCCGC 837
         ||||||||||||||:::      :::|||||||::
Db   239 AlaAlaGluGluLysSerValHisSerThrProGluGlyCysSerSerAlaCysLys 258

Qy   838 ATGATTCTTGAAATCATGCAGAAGAGGCAGCAGAGAGACCAAACTAGCCGAAGAGATTCCT 897
         ||||||||||||||||||||||||   :::|||::
Db   259 MetIleLeuGluIleMetHisLysGluGluMetHisLysLysValThrAlaAspValPro 278

Qy   898 CTGAAAATCTTGGCCACCAACATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAAT 957
         ||||||||||   :::|||||||||||      |||||||
Db   279 LeuLysIleLeuAlaHisAsnAlaAsnPheValGlyArgLeuIleGluGlyLysGluGlyAlaAsn 298

Qy   958 TTGAAGAAAATTGAACATGAAACAGGACCCAAGATAACAATCCTCATCTTTGCAGGATTTG 1017
         ||||||      ::||||      |||||||||||||||
Db   299 LeuLysLysValGluGlnLeuAspThrGluThrIleSerGluSerLeuGlnAlaAspLeu 318

Qy  1018 AGCATATACACCCGGAAAGAACCATCACTGTGAACGGCACAGTTGAGCGCTGTGCCAGT 1077
         ::::|||||||||||||||||||:::||||
Db   319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIleGluAsnCysCysSerArg 338

Qy  1078 GCTGAGATAGAGATTATGAAGCAGTGCGTGAGGGCCTTTGAAAATGATATGCTGGCTGTT 1137
         ||||||:::||||||||||||||||||
Db   339 AlaGluGluGlnIleGluIleMetLysGlnCysValArgGluAlaIleTyrGluAlaAspProValAlaAlaAlaMet 358
```

```
Qy  1138 AACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGGCATCTTTCAACA 1197
         :::      ||||:::||      ||||      ::::||
Db   359 SerLeuGlnSerHisLeuIlePheProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAla 378

Qy  1198 GGACTGTCCGTGCTATCTCACCAGCAGGGCCCGGGAGCTCCCCGCTGCCCCCTAC 1257
         |||      ::||||   ||   -----ProSerValThrGlyAlaAlaProTyr
Db   379 SerSerSerAlaValProPro----- 396

Qy  1258 CACCCCTTCACTACCCACTCCGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGC 1317
         ||||||||||||||
Db   397 SerSerPheMet------------------------------------------- 400

Qy  1318 CCGTTCCCGCCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACC 1377
         |||||::||||   :::|||||
Db   401 ------------------GlnAlaProGluGlnGlnGlyMetValGlnPheIleProAla 414

Qy  1378 CAGGCTGTGGGGGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGCCGAGATTC 1437
         ||||||||||||||:::|||||||||
Db   415 GlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGluSerArgPhe 434

Qy  1438 GCCGGAGCCTCTATCAAGATTGCCCTGCCGGAGCAGGCCACAGACTTCAGCGAAACGGATGGTC 1497
         |||||||||||||      ||||||||
Db   435 AlaSerAlaSerIleLysIleAlaLeuProProGluThrProAspSerLysValArgMetVal 454

Qy  1498 ATCATCACCGGGCACCGGGCCAGTTCAAGGCCCAGGGACGACGGATCTTTGGGAAACTTG 1557
         ::||||||||||||||||||||      :::|||||
Db   455 ValIleThrGlyProGlyProGluGlnGlnLysValAlaGlnGlnLysArgIleTyrGlyLysLysLeu 474

Qy  1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCTGAAGAGGTGAGC 1617
         ||||||||||||:::|||||||||
Db   475 LysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIleArgVal 494

Qy  1618 CCCCTTCCACAGCTGGCGCGGGTGATTGGCAAAGGTGGAGCAAGACCGTGACAAGACTGCAG 1677
         |||||||||||||      :::||||
Db   495 ProAlaSerAlaAlaGlyAlaGlyYrAlaValIleGlyLysGlyLysIleThrValAsnGluLeuGln 514

Qy  1678 AACTTAACCAGTGCAGAGTGATCGTGCCTCGTGACCAAACGCCAGATGAAATGAGGAA 1737
         |||:::      ||||||||||:::|||||
Db   515 AsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsnAspGln 534

Qy  1738 GTGATCGTCAGAATTATCCGGCACTTCTTTGCTAGCCAGACTGCACAGCCAAGAATCACAGG 1797
         |||||||||      :::|||||||
Db   535 ValIleValLysIleIleGlyHisAsnPheValAlaSerGlnMetAlaGlnArgLysIleArg 554

Qy  1798 GAAATTGTACAACAGGTGAAGCAGCAGAGAGCCTCAGGGAGTCGCCTCACAG 1857
         :::||::||      ||||||
Db   555 AspIleLeuAlaGlnIleValLysGlnValGlnHisGlyGlnSerAsnLeuAlaGlnAlA 574

Qy  1858 CGGAGCAAG 1866
         ||||||
Db   575 ArgArgLys 577
```

RESULT 11

Q9PW80

ID   Q9PW80;   PRELIMINARY;   PRT;   582 AA.
AC   Q9PW80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vg1 RNA binding protein (Decapentaplegic and Vg-related 1, RNA binding
DE   protein).
GN   DVR1RBP.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA   Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.;
RT   "Vg1 RBP intracellular distribution and evolutionarily conserved

RT   expression suggest multiple roles during development.";
RL   Mech. Dev. 0:0-0(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB; TISSUE=Body;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB; TISSUE=Body;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF161270; AAD45610.1; -.
DR   EMBL; BC045873; AAH45873.1; -.
DR   ZFIN; ZDB-GENE-000308-1; dvr1rbp.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE 582 AA; 63351 MW; 9DAE63200681B306 CRC64;

Alignment Scores:

Pred. No.:			
Score:	2.29e-138	Length:	582
	1976.00	Matches:	393
Percent Similarity:	79.19%	Conservative:	75
Best Local Similarity:	66.50%	Mismatches:	101
Query Match:	32.28%	Indels:	22
DB:	13	Gaps:	4

US-09-270-437D-6 (1-3412) x Q9PW80 (1-582)

```
Qy    313 AGGAAAATTCAGATTCGAAACATCCCTCCTCGAGTGGAGGTGTTGATGGACTT 372
              ||:||||||||||||||||||||||||||
Db     81 CysLysLeuGlnIleArgAsnIleProProHisMetGlnTrpGluValLeuAspGlyLeu 100

Qy    373 TTGGCTCAATATGGACAGTGGAGAATCTGGACAACAGTCAACACAGACACAGAAACCGCC 432
              ||||||||||:::
Db    101 LeuAlaGlnTyrGlyThrValGlyGlnSerCysGluValAsnThrAspThrGluThrAla 120

Qy    433 GTTGTCAACGTCACATATGCAACAAGAAGAAGACCATGGAAGAAGCTAAGC 492
              ||||||||||||:::::::
Db    121 ValValAsnValArgTyrGlyAlaLysAspAlaArgGluAlaMetAspLysLeuAsn 140

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCATCCGGATGAAGAGGTGAGC 552
              |||      ||||:::::
Db    141 GlyPheLeuMetGluAsnTyrArgAlaLeuLysProGluThrAlaAla 160

Qy    553 TCCCCTTCGCCCCCTCAGCGAGCGCCAGGCTGGG------GACCACTCTTCCGGAGCAA 606
              :::  ||:::||
Db    161 AlaAspAlaProAlaValGlyValGlyArgAlaGlyPheAsnProArgGlyProProArgGln 180

Qy    607 GGGCCACGCCCCTGGGGGCACTTCTCAGGCCACAGCAGATTGATTCCCGCTGCGGATCCTG 666
              |||
Db    181 GlySerProSerLeuGlyAlaArgProLysLeuGlnSerAspValProLeuArgLeuLeu 200

Qy    667 GTCCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGACATC 726
              ||||||||||||||||||||||||
Db    201 ValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyLysAlaThrIleArgAsnIle 220

Qy    727 ACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAAACTCTGAGCTGCAGAG 786
              ||||||||||||||||
Db    221 ThrLysGlnThrHisSerLysIleAspIleHisArgLysGluAsnAlaGlyAlaAlaGlu 240

Qy    787 AAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATGAGCATTCTT 846
              ||||||:::::
Db    241 LysProIleThrValHisSerThrProGluGlyLeuPheValGlyCysSerSerArgAsnIleMet 260

Qy    847 GAAATCATGCAGAAGAGGCAGATGAGACAAACTAGCCAGAAGAGATTCCTCTGAAAAATC 906
              |||||||||||||||
Db    261 GluIleMetGlnLysGlnArgGlnThrLysLeuAlaIleAspThrLysIleIleProLeuLysLysIle 280

Qy    907 TTGGCACCAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAGGCAGAAATTTGAAGAAA 966
              ||||||||||||:::::
Db    281 LeuAlaHisAsnAsnPheValGlyArgLeuIleGlyArgAsnLeuLysLys 300

Qy    967 ATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATAC 1026
              ::||||||||||::::::
Db    301 IleGluGlnAspThrArgThrLysIleThrIleSerProLeuGlnAspLeuThrLeuTyr 320

Qy   1027 AACCCGGAAGAACCATCACTGTGAAGGGCCACAGTTGAGCGCCTGCCAGTGCTGAGATA 1086
              ||||||||||||||||||
Db    321 AsnProGluArgThrIleThrValLysGlyThrLeuAspAlaCysAlaLysAlaGluGlu 340

Qy   1087 GAGATTATGAGAAGCTGCGTGAGCGTTTGAAAATGATATGCGGCTGTTAACCAACAA 1146
              ||||||:::::
Db    341 GluIleMetLysLysValArgGluSerTyrGluAsnAspValAlaAlaMetHisLeuGln 360

Qy   1147 GCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTGTCC 1206
              ||||:::::
Db    361 SerAsnLeuIleProGlyLeuGluAsnLeuAsnAlaLeuGlyLeuPheProGlyAlaAlaSer 380

Qy   1207 GTGCTATCTCCACCAGGAAGCTGCCCCGCGGAGCTCCCCGCTGCCCCCTACCACCCCTTC 1266
              :::
Db    381 GlyGlyIleSerProSerValValSerGlyProProProGlyAla----- 395

Qy   1267 ACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCGAGACCCCTTCTCGGGTAAA 1326
              :::||
Db    396 ------GlnAlaGlyTyrGlnSer------ 403

Qy   1327 CATCATCACTCTTATCCAGACGCAGGAGATTGAATCTCTTCATCCCAACCCAGGCTGTG 1386
              :::
Db    404 ------AlaGlnMetGluSerGluThrValHisLeuPheIleProAlaLeuAlaVal 420

Qy   1387 GGGCGCCATCATCGGGAAGAAGAGAGGGGCACACATCAAACAGTCTGGCGAGATTCGCGGAGCC 1446
```

```
Qy     73 ATGAACAAGCTTTACATGGGACCTGGAACCTCAGAGCCCGTCACCGTCACCGGACCTTCCGGCAG 132
              ||||||||:||||||:||||::||  ||||||
Db      1 MetAsnLysLeuTyrIleGlyAsnValSerGluGlnAlaSerAlaLeuAspLeuGluSer 20

Qy    133 CTCTTTGGGGACAGGAAGCTGCCCCTGCCGGGGACCAGGTCCTGCTGAAGTCCGGCTACGCC 192
              ::||:||::|||   |||  :::|||
Db     21 IlePheGluGlnTrpLysIleProPheSerAlaProPheLeuValLysSerGlyTyrAla 40

Qy    193 TTCGTGGACTACCCCGACCAGAACTGGGCCCATCCGGCCATCCGCGCCATCGAGACCCCTTCTCGGGTAAA 252
              |||||||:|||:::    |||  :::||||||||||    ||||
Db     41 PheValAspCysProAspGluLysProLysValAlaAlaMetArgAlaIleAspThrLeuSerGlyGlyLys 60

Qy    253 GTGGAATTGCATGGGAAATCATCATGGAAGTTGATTACTCAGTGTCTAAAAGCTAAGGAGAGC 312
              |||||:|||:|:|||||  |||:|||:|||||||||
Db     61 ValGluLeuHisGlyLysValLeuGluGlnValHisSerValProLysSerValProLysArgSer 80
```

```
Db        ||||||||||||||||||||        ::::|||       ||||||||||||||||||||||||
      421 GlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 440

Qy   1447 TCTATCAAGATTGCCCTGCGGAAGGCCCAGACGTCAGGCGAAAGGATGGTCATCATCACC 1506
          ||||||||||||||||||        ::::|||
Db    441 SerIleLysIleAlaProAlaAspGlyIleAspAlaLysGlnArgMetValIleIleSer 460

Qy   1507 GGGGCACCGGAAGCCCAGTTCAAGGGACGGACGGATCTTTGGGAAACTGAAAGAGAGAA 1566
          ||||||||||||||||||||||||||||||||||||||
Db    461 GlyProProGluAlaGlnPheLysAlaGlnGlyLyArgIlePheGlyLysLeuLysGluGlu 480

Qy   1567 AACTTCTTTAACCCAAAGAAGAAGAGAGTGAAGCTGGAGAGAGAGCATATCAGAGTGCCCTCTTCC 1626
          |||||||||||||||||||||||||||||||||||||::::|||||
Db    481 AsnPhePheGlyProLysGluGluLysLeuGluAlaHisIleLysValProSerPhe 500

Qy   1627 ACAGCTGGCGCGGGTGATTGGCAAAGTGCCAAGACCGTGAACGAACCTCAGAACTTAACC 1686
          ||||||||||||||||||||||||||||||||::::||||||||||
Db    501 AlaAlaGlyAlaArgValIleGlyLysValProLysThrValAsnGluLeuGlnAsnLeuThr 520

Qy   1687 AGTGCAGAAGTCATCGTGCTCGTGACCAAACGCCAGATGAAAATGAGAGAGAGTGATCGTC 1746
          ||||||||||||||||||||||||||::::::|||||:::||||
Db    521 SerAlaGluValValValProArgAspGlnThrProAspGluLysMetArgGluSerLeu 540

Qy   1747 AGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGGCCAAGATCAGGGAAATTGTA 1806
          :::||||||||||||||||::::||||||||||||::::|||||::||||||||
Db    541 LysIleThrGlyHisPhePheTyrAlaSerGlnThrAlaGlnAlaLysIleArgGluIleIle 560

Qy   1807 CAACAGGTGAAGCAGCAGCAGAGCAGCAGAAATACCCT 1839
          ||:::::||||||||||:::::|||||||
Db    561 SerGlnValArgArgArgGlnGlnProLysPro 571


RESULT 12
Q8C2J9
ID   Q8C2J9;     PRELIMINARY;       PRT;    579 AA.
AC   Q8C2J9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Insulin-like growth factor 2.
GN   IGF2BP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK088465; BAC40370.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
GO   GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;   63551 MW;   937B601A95D06B77 CRC64;


Alignment Scores:
Pred. No.:
Score:                    2.31e-137
Percent Similarity:       1962.50
Best Local Similarity:    79.39%
Query Match:              66.05%
                          32.06%

Length:          579
Matches:         391
Conservative:    79
Mismatches:      91
Indels:          31
```

```
DB:                                           11                        Gaps:                 8

Q8C2J9
US-09-270-437D-6 (1-3412) x Q8C2J9 (1-579)

Qy     73 ATGAACAAGCTTTACATCGGAACCTGAGCCCGTCACCGCCAGCGACCTCCGGCAG 132
          ||||                      |||||
Db      1 MetAsnLysLeuTyrIleGlyAsnLeuSerArgHisAlaGlyProLeuAlaGluSer 20

Qy    133 CTCTTTGGGGACAGGAGGAAGCTGCCCTGGCGGACAGCAGGTCCTGTGAAGTCCGGCTACGCC 192
          :::||||            |||||||||||
Db     21 ValPheLysAspAlaLysIleProValAlaGlyProPheLeuValLysThrGlyTyrAla 40

Qy    193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGGCGCCATGAGACCCTCCGGGTAAA 252
          |||||       ||||||||        ||||||
Db     41 PheValAspCysProAspGluGlyTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60

Qy    253 GTGGAATTGCATGGGAAAATCATGAAGTTGATTACTCAGTCTCTAAAAGTTAAGGAGC 312
          :::|||::::|||:::|||
Db     61 MetGluLeuHisGlyLysProMetGluValGluHisSerValProLysArgGlnArgIle 80

Qy    313 AGGAAAAATTCAGATTCGAAACATCCCTCTGCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
          |||
Db     81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy    373 TTGGCTCAATATGGACAGTGAGAATGTGGAACAAGTCAACACAGACAGAAAACCGCC 432
          |||||:::
Db    101 LeuValGlnTyrGlyValValGluSerGlyGlnValAsnThrAspSerGluThrAla 120

Qy    433 GTTGTCAACGTCACATATGCAACAAGAAGAAGAGAGCAAAAATACCATGGAGAAGGAGCTAAGC 492
          ||||||||:::::|||||::::|||
Db    121 ValValAsnValThrTyrSerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsn 140

Qy    493 GGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAGC 552
          ||||          |||||::::        ::
Db    141 GlyPheGlnLeuGluGluAsnPheThrLeuLysValAlaTyrIleProMetLysArgAlaAla 160

Qy    553 ------TCCCCTCGCCCCCTCAGCGAGCCCAGCGTGGGGAC--------CACTCTCC 597
          ||||
Db    161 GlnGlnAsnProSerProGlnLeuArgGlyArgArgGlyValProGlyGlnThrGlySerSer 180

Qy    598 CGGGAGCAAGGCCACGCCCCTGGGGCACTCTCAGGCCAGACAGATGATTTCCGCGCTG 657
          |||:::                   :::  :::||||
Db    181 ArgGlnAla------SerProGlySerValSerGlnLysGlnLysProCysAspLeuProLeu 198

Qy    658 CGGATCCTCGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAGGAGGGCTTGACCATA 717
          ||:::||||||::::        |||||||
Db    199 ArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysLysGluValAlaThrIle 218

Qy    718 AAGAACATCACTAAGCAGACCAGTCCCGGGTAGATATCCAGGTAGAAAAGAGAACTCTGGA 777
          ||||||||||||        ||||||||:::||||
Db    219 ArgAsnIleThrLysGlnThrGlnSerLysAspValHisArgLysGluAsnThrGly 238

Qy    778 GCTCGAGAGAGCCTGTCACCATCCATGCAGCCCCAGAGGGGACTTCTGAAGCATGCCGC 837
          |||||||                  :::  :::||||
Db    239 AlaAlaGluGluGlySerIleThrIleLeuSerThrProGlyThrSerAlaAlaCysLys 258

Qy    838 ATGATTCTTGAAATCATCGAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCT 897
          |||||||                  :::   :::||||
Db    259 SerIleLeuGluIleMetHisLysGluAlaGlnAsnIleLysPheThrGluIleIlePro 278

Qy    898 CTGAAAAATCTGGCACACAGGCTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAAT 957
          ||||||||::::||||||||
Db    279 LeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsn 298

Qy    958 TTGAAGAAATTGAACATGAAACACGGACCAAGATAACAATCCATCTTTCAGGATTTG 1017
          |||||||||||||
Db    299 LeuLysLysIleGluLeuAspThrArgAspThrLysIleThrIleSerProLeuGlnGluLeu 318

Qy   1018 AGCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGT 1077
          :::::||||||||
Db    319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlySerValGlnThrCysAlaLys 338

Qy   1078 GCTGAGATAGAGATTATGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGCTGTT 1137
```

```
Db   339  AlaGluGluGlyIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMet  358
               ||||||     |||||||||     ::::|||||||||::||||||      :::::
Qy  1138  AACCAACAGCCAATCTGATCCCAGGGTTGAACTCAGGGCACTTGGCATCTTTTCAACA  1197

Db   359  AsnLeuGlnAlaAsnLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe-----  376
               ||||         |||                             |:::|||
Qy  1198  GGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGGAGCTCCCCCGTGCCCCCTAC  1257

Db   377  ----------------ProProThrSer------GlyMetProProProThr------  386
                          |||||        :::||||
Qy  1258  CACCCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGC  1317

Db   387  ----------SerGlyProSerThrLeuThrPro----------  395
                  ||:::|||    |||
Qy  1318  CCGTTCCGCATCATCACTCTTATCCAGACAGGAGATTGAATCTTCATCCCAACC  1377

Db   396  ProTyrProGlnPheGluGln--SerGluThrGluValHisLeuPheIleProAla  414
               ||:::|||
Qy  1378  CAGGCTGTGGGGCGCCATCATCGGGAGAGAAGGGGGCACACATCAAACAGCTGGGGAGATTC  1437

Db   415  LeuSerValGlyAlaIleIleGlyIleGlyGlnGlyIleHisIleLysGlnLeuSerArgPhe  434
               :::||||||||||||||||||:::|||||||||||||:::|||||||||:::|||
Qy  1438  GCCGGAGCCTCTATCAAGATTGCCCTGGGAAGGCCCAGACGTCAGCGAAAGGATGGTC  1497

Db   435  AlaGlyAlaSerIleLysIleAlaLeuAlaProAlaGluAlaProAlaLysValArgMetVal  454
               ||:::|||
Qy  1498  ATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACCGATCTTTGGGAAACTG  1557

Db   455  IleIleThrGlyProProGluAlaGlnPheLysAlaGlnAspArgIleTyrGlyLysIle  474
               |||||||||||||:::||||||||||||||||||||||||||:::|||
Qy  1558  AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGCTGGAAGCCGCATATCAGAGTG  1617

Db   475  LysGluGluAsnPheValSerProLysSerGluGluGluLeuGluAlaHisIleArgVal  494
               ||||         :::|||||:::||||||||||||||||||||:::|||
Qy  1618  CCCTCTTCCACAGCTGGCCGGGTGATTGGCGCAAAGGTGGCAAGACCGTGAACGAACTGCAG  1677

Db   495  ProSerPheAlaAlaGlyArgValIleGlyLysLysGlyLysThrValAsnGluLeuGln  514
               ||:::|||||||||||||||||||||||||||||||||||||||||:::|||
Qy  1678  AACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAA  1737

Db   515  SerLeuSerSerAlaGluValValProArgAspGlnAsnAspGln  534
               :::||:::|||||||||||||||||||||:::|||
Qy  1738  GTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAAGCTCACAGCGCCAAGATCAGG  1797

Db   535  ValValValLysIleThrGlyHisHisPheTyrAlaCysHisGlnValAlaGlnArgLysIleGln  554
               |:::||:::|||||:::|||
Qy  1798  GAAATTGTACAACAGGTGAAGCAGCAGGAGCAGAAA  1833

Db   555  GluIleLeuThrGlnInValLysGlnHisGlnGlnGln  566
```

RESULT 13

ID   Q9CPN8   PRELIMINARY;   PRT;   579 AA.
AC   Q9CPN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein 3)
DE   (Insulin-like growth factor 2, binding protein 3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;

RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.

CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   EMBL; BC045138; AAH45138.1; -.
DR   EMBL; BC049082; AAH49082.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA; 63574 MW; CABD9A4355B392B7 CRC64;

Alignment Scores:
Pred. No.:   5.45e-137   Length: 579

Score: 1957.50   Matches: 390
Percent Similarity: 79.39%   Conservative: 80
Best Local Similarity: 65.88%   Mismatches: 91
Query Match: 31.98%   Indels: 31
DB: 11   Gaps: 8

US-09-270-437D-6 (1-3412) x Q9CPN8 (1-579)

```
Qy     73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCCGTCACCGGCGACCTCCGGCAG 132
          |||||||||||||||||||||||||| :|||||||||||  ::|| |||||||||
Db      1 MetAsnLysLeuTyrIleGlyAsnLeuSerAspHisAlaGlyProAlaAspLeuGluSer 20

Qy    133 CTCTTGGGGACAGGAAGCTGCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACCGC 192
          ::|| ||  |||||::|||  ||:|| ||||| ||:::||||||||||||||
Db     21 ValPheLysAspAlaLysIleProValAlaGlyProPheLeuValLysThrGlyTyrAla 40

Qy    193 TTCGTGGACTACCCGACCAGAACTGGGCCATCCGCGACCCCTCTGGGTAAA 252
          |||||||| |||||||||||||||||||||||||||||||  ::|| :|||
Db     41 PheValAspCysProAspGlyGlyTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLys 60

Qy    253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
          :::||||||| ||||||||||||||||||||||::||||| |||::||||
Db     61 MetGluLeuHisGlyLysProMetGluValGluHisSerValProLysArgGlnArgIle 80

Qy    313 AGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGATGGACTT 372
          :::||:::||||||||||| ||||||||||||||||| ||||
Db     81 ArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy    373 TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAAACCGCC 432
          ||||||||||||||||||||||||||||| |||||||||::||||||||
Db    101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSerGluThrAla 120

Qy    433 GTTGTCAACGTCACATATGCAACAGAGAAGAGCCATGAGAAGCTAAGC 492
          |||||||||||:||||||||||:::|||||||||||| |||::|| |||
Db    121 ValValAsnValThrTyrSerAspLysSerGluTyrSerArgGlnLeuAsn 140

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC 552
          ||||||||||||||||| :::||||||||||||||  ||||||||||||::||
Db    141 GlyPheGlnLeuGluAsnPheLysLysValAlaTyrIleProGluProGluThrAlaAla 160

Qy    553 ------TCCCCTGCCCCCCTCAGCGAGCCCAGCCTGGGGAC--------CACTCTTCC 597
          :::||||||||||||| |:||||||||||||       |||||||
Db    161 GlnIleAsnProSerProGlnLeuArgGlyArgArgGlyPheProGlyGlyArgGlySer 180

Qy    598 CGGGAGCAAGGCCACGCCCTGGGGCACTTCCAGGCCACAGATTCGATTTCCCGCTG 657
          |||::    ::||||   ||||||||||||||||||||||::|||  |||||
Db    181 ArgGlnAla------SerProGlySerValSerLysProLysSerProCysAspLeuPro 198

Qy    658 CGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGAAAGGAGGGCTTGACCATA 717
          |||||||||||||||||||||||||||||||| |||::|||||||||||||||||||
Db    199 ArgLeuLeuValProThrGlnPheValGlyAlaIleIleGluLysGluGlyLeuAlaThrIle 218

Qy    718 AAGACACTACTAAGCACCATTGCTGTCCCGGGTAGATATCCATAGAAAGAGAACTCTGA 777
          :::|||||||||||||||||||||:::|||||||||::::|||:::|| :::
Db    219 ArgAsnIleThrLysGlnInThrGlnSerLysIleSerAsnValHisSerArgLysValAsnThrGly 238

Qy    778 GCTCGAGAGAAGCCTGTCACCATGCAGAGGGGACTTCTGAAGCATGCCGC 837
          |||:::||||||||| :::|| ||||| :|||||||||::||||||:::
Db    239 AlaAlaGlyLysSerIleThrIleLeuSerThrProGluGlyThrSerArgAlaAlaCysLys 258

Qy    838 ATGATTCTTGAAATCCATGACGAAAAGAGAGGCAGATGAGACCAAAACTAGCCGAAGATTCCT 897
          |||||||:::|| ||||  :|:: ||||||||||||||||||::||
Db    259 SerIleLeuGluIleHisMetHisLysLeuGlnAlaGlnAspIleLysPheThrGluIlePro 278

Qy    898 CTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGGCAGAAAT 957
          ||||||||:::|||||||:::||||||||||||||||||||||||
Db    279 LeuLysIleLeuGluAlaHisAsnAsnPheValGlyArgLeuLeuGlyLysGluGlyAlaAsn 298

Qy    958 TTGAAGAAAATTGACATGAAACAGGACCAAGATAACAATCCATCTTTGCAGGATTTG 1017
          ||||||||:::|||||||||::||||||||:||| ||||||||::|||
Db    299 LeuLysLysIleGluGluAspAlaAsnThrArgSerIleThrLysSerProLeuGlnGluLeu 318

Qy   1018 AGCATATACAACCCGGAAAAGAACCATCACTGTGAGGGCACAGTTGAGGGCCTGCCAGT 1077
```

```
Db    319 ThrLeuTyrAsnProGluArgThrIleThrValLysGlySerValGluThrCysAlaLys 338

Qy   1078 GCTGAGATAGAGATTATGAAGAGAGCTGCGTGAGGCCTTGAAAATGATGATGTGGCTGTT 1137
          ||||||| :::||| |||||:::|||| ::: ::: :::
Db    339 AlaGluGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMet 358

Qy   1138 AACCAACAAGCCAATTGATCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACA 1197
          ||||||||||||||::|| ||||||||||||||||||||:::|||||||||
Db    359 AsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe------ 376

Qy   1198 GGACTGTCCGTGCTATCTCCACCAGGCAGCAGGGCCCCGGAGCTCCCCCGCTGCCCCCCTAC 1257
          ||||||||| |||||||||||||||| ||::||||||||
Db    377 ------ProProThrSer-----GlyMetProProProThr------ 386

Qy   1258 CACCCCTTCACTACCCACTCCGATACTTCTCCAGCTGTACCCCCATCACCAGTTTGGC 1317
          ||||||||| |||||::|||| :::||||  ||||
Db    387 ------SerGlyProSerThrLeuThrPro------ 395

Qy   1318 CCGTTCCCGCCATCATCCTCTATCCAGACGACGAGATTGTGAATCTTCATCCACC 1377
          |||::||||||||||||||||||||||||||||||||::||||||||||||||
Db    396 ProTyrProGlnPheGluGln---SerGluThrGluValHisLeuPheIleProAla 414

Qy   1378 CAGGCTGTGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGGAGATTC 1437
          ||||||||||||||||||||||||||::|||||||||||||||||||
Db    415 LeuSerValGlyAlaIleIleGluIleGlyLysGlnLeuLysIleLeuLysSerArgPhe 434

Qy   1438 GCCGGAGCCTCATCAAGATTGCCCCTGCCGAGGCCCAGACGTCAGCGAAGGATGGTC 1497
          ||||||||||||||||||| ||::||||||::|||||||||||||||
Db    435 AlaGlyAlaAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetVal 454

Qy   1498 ATCATCACCGGCCACGTGGCCGCGGGTGATTGGCAAGAAGTGGCAAGAACGAACTGCAG 1557
          |||||||||||||||::|||||||||||||||||||||||||||||::||
Db    455 IleIleThrGlyHisGluProGlnLeuAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLysIle 474

Qy   1558 AAAGAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGGCCATATCAGAGTG 1617
          |||||||||||||||::|| :::||||||||||||||||||||||||||||
Db    475 LysGluGluAsnPheValSerProLysGluGluSerGluAlaGlyArgProTyrGlnSer 494

Qy   1618 CCCTCTTCCACAGCTGGCCGGGGTGATTGGCAAGAAGTGGCAAGACCGTGAACGAACTGCAG 1677
          |||::|||||||||::||:::||||||||||||::|| :::::|:::
Db    495 ProSerPheAlaAlaGlyLyArgValIleGluGlyLyGlyLysThrValAsnGluLeuGln 514

Qy   1678 AACTTAACCAGTGCAGAGTCATCGTGCCTGTTCGTGACCAAACGCCAGATGAAAATGAGGA 1737
          |||::||:::|||||||||:::||:::|||||||||::|||||:::
Db    515 SerLeuSerSerAlaGluValValProArgAspGlnThrProAspGluAsnAspGln 534

Qy   1738 GTGATCGTCAGAATTATCGGGCACTTCTTGTTGTAGCCAAGATCAGG 1797
          |||||| :::|:::|||||||||||||||||||||||||||||||||
Db    535 ValValValLysIleThrGlyHisPheTyrAlaLysGlnValAlaArgLysIleGln 554

Qy   1798 GAAATTGTACAACAGGTGAAGCAGCCAGGAGCAGAAA 1833
          ||||||||||||||::|| :::||:::|:::
Db    555 GluIleLeuThrGlnValLysGlnHisGlnGlnGln 566
```

RESULT 14
00425

ID   00425   PRELIMINARY;   PRT;   579 AA.
AC   00425;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative RNA binding protein KOC (KOC).
GN   KOC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mueller-Pillasch F., Lacher U., Wallrapp C., Et Al.;
RL   Oncogene 0:0-0(0).
RN   [2]

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; U97188; AAC35208.1; -.
DR   EMBL; U76705; AAD09223.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0006412; P:protein biosynthesis; TAS.
DR   GO; GO:0006396; P:RNA processing; TAS.
DR   InterPro; IPR004087; KH dom.
DR   InterPro; IPR004088; KH type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH TYPE 1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE 579 AA; 63720 MW; AE5C3AEE3C135C5 CRC64;

Alignment Scores:
Pred. No.:                                      579
Score:                1.81e-136    Length:        579
Percent Similarity:   79.02%       Matches:       386
Best Local Similarity: 65.31%      Conservative:   81
Query Match:          31.87%       Mismatches:     95
DB:                   4            Indels:         29
                                   Gaps:            8

US-09-270-437D-6 (1-3412) x O00425 (1-579)
```

```
Qy    73 ATGAACAAGCTTTACATCGGGAACCTGGGACCCCCGTCACCGGCGAGACCTCCGGCAG 132
          |||||||||||||||||||||||                             ||||||
Db     1 MetAsnLysLeuTyrIleGlyAsnLeuGlyProSerGluAsnAlaAlaProSerLeuGluSer 20

Qy   133 CTCTTTGGGGACAGGAAGTGCCCCTGGGGGACCAGTTCCTGCTGAGTCCGGCTACGCC 192
          :::|||   ||| ||  :|:   |||||::|||||:||||| || :|:||:||:||||
Db    21 IlePheLysAspAlaLysIleLeuProValSerGlyProPheLeuValLysThrGlyTyrArgAla 40

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCGCCATGAGAACCCTCTCGGGTAAA 252
          |||||||||   |||:||:||  ||||:::|||||||||   ::||||||:|||||||
Db    41 PheValAspCysProAspGluSerTrpAlaLeuLysLysAlaIleGluAlaLeuSerGlyLys 60

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTAAGGAGC 312
          |||||||:||||||||  :||||:|||||:|||:||::| :||:||||| ||
Db    61 IleGluLeuHisGlyLysProIleGluValGluHisSerValProLysArgGlnArgIle 80

Qy   313 AGGAAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTT 372
          |||||||||||||||||:::|||| ||:|||||:||:|||||||||||:|||||||||||
Db    81 ArgLysLeuGlnIleLeuArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100

Qy   373 TTGGCTCAATATGGACAGTGGAGAATGGGAGAAGGAAGTGGAACAAGTCAACACAGAAACCGCC 432
          ||||||||||||||| ||:|||::|  ::|||:|||||:|||||||:||||||||||| |||
Db   101 LeuValGlnTyrGlyValValGluSerCysGluGluValGluLysAsnThrAspSerGluThrAla 120

Qy   433 GTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAATAGCCATGGAGAAGCTAAGC 492
          |||||| ||::|::||||||  :||::||||:|| |||::|||||||||||||||||:
Db   121 ValValAsnValThrTyrSerGluLysSerAspGlnAlaArgGlnAlaArgGlnAlaLeuLeuAspSerLeu 140

Qy   493 GGGCATCGAGTTGTGAGAACTACTCCTTCAAGATTCCTACATCCCGATGAGGAGGTGAGC 552
          ||||||:|||||:||:|||| ||:|||:|||||||||||| |||:|||||:|||:|||
Db   141 GlyPheGluLeuGluAsnPheValSerValAlaTyrIleProAspGluMetAlaAla 160

Qy   553 TCCCCTTCGCCCCCCTCAG------CGAGCCCAGCGT------GGGGACCACTCTTCCCGG 600
          |||||||:||||| |:: |      |||  :|:||:|      ||||||:||:|||||||
Db   161 GlnGlnAsnProLeuGlnProLeuArgGlyLeuArgGlyLeuGlnArgGlySerSer 180

Qy   601 GAGCAAGGCCACCGGGGGACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
          |||||||||:|||| ::|||||:|| ::|||||||||||| |:|:  :|:||||
Db   181 ArgGlnGly---SerProGlySerValSerLysGlnGlnLysProCysAspLeuProLeuArg 199
```

```
Qy   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATGGAAAGGAGGGCTTGACCATAAAG 720
          :::|||||||||||||||||||||||| ||||||||||||||| |||  :|:   :::
Db   200 LeuLeuValProThrGlnPheValGlyAlaIleIleGluLysGlyAlaThrIleArg 219

Qy   721 AACATCACTAAGCAGACAGACCCGGTAGATATCATCAGAGAAAGAGAACTCTGGAGCT 780
          |||||||:||||||| || ||:|||||:|||||||::|||:||:||::|||||:|||
Db   220 AsnIleThrLysGlnThrArgGlnSerLysIleIleAspValHisArgLysGluAsnAlaGlyAla 239

Qy   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATG 840
          |||||||||||||||||||||||||||||||||:||||||||||||||||||||:||
Db   240 AlaGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSer 259

Qy   841 ATTCTTGAAATCATGCAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
          |||||||||:||||:|::|||||:||| ||||||||||||||||||||||||||||||
Db   260 IleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGlnIleProLeu 279

Qy   901 AAAATCTTGGCACACAATGGCTTGGTTGGAGACTGGAAAGAAGGCAGAAATTG 960
          |||||||||:|||||||||||||||||:|||||||||||:|||||||||||||
Db   280 LysIleLeuAlaHisAsnAsnPheValGlyGlyAspTrpLysGluGlyArgAsnLeu 299

Qy   961 AAGAAAATTGACATAAAGAGGACCAAGAATAACAATCCATCTTTGCAGGATTTGAGC 1020
          |||||:||:|||||:|||:||:|||||||  :||  |||||:|:||||||||||||
Db   300 LysLysIleGluGlnIleAspThrArgThrLysThrIleThrIleSerProLeuGlnAspLeuThr 319

Qy  1021 ATATACACCGGAGAAACCATCACTGTGAAGGGCACAGTTGAGCGCTGTGCCAGTGCT 1080
          |:|||||||||:|||||||||||:|||||||||||||||||:|||:||:|||||:|||
Db   320 LeuTyrAsnProGluArgThrIleThrValLysGlyAsnValGlyThrCysAlaLysAla 339

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCTGAGGCTTTGAAAATGATAATGTGGCTGTTAAC 1140
          |||||:|||||||||||||||||:||:|||||:  ::|  :::: |||||||:||||
Db   340 GluGluGluIleMetLysLysLeuLeuArgGluSerTyrGluAsnAspIleAlaSerMetAsn 359

Qy  1141 CAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCCACTTGGCATCTTTTCAACAGGA 1200
          |||||||||||:|| ::||||||:|||||:||||||||||||:||:||:|:-------
Db   360 LeuGlnAlaHisLeuIleLeuProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe-------- 376

Qy  1201 CTTGTCCCGTCTATCTCCACCAGCAGGGCCCCGGAGCTCCCCCGGGAGCTCCCCCTACCAC 1260
          |||||:|||||:|| |:|:::|:::||:|||-------GlyMetProProPro------------
Db   377 ----------------------GlyMetProProPro----------ProProThr----- 386

Qy  1261 CCCTTCACTACCCCCGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCCG 1320
          |||||||||||:|||:|||||||:|||||||-----------Pro--------
Db   387 ----------SerGlyTyrProProSerAlaMetThrPro-------- 396

Qy  1321 TTTCCCGCATCATCACTCCTATCCAGCAGGAGATTGTGAATCTTCATCCCAACCCAG 1380
          |||||:||||:|||||||||||||||||||:|||||:|||||:|||||||||||||
Db   397 TyrProGlnPheGluGln---SerGluThrValHisSerGlnPheLeuProAlaLeu 415

Qy  1381 GCTGTGGGGCCCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCC 1440
          ||:||:|||:||||||||:|||||||||||:|||||||||||||||:|||:|:|||||:|||
Db   416 SerValGlyAlaIleIleGlyLysLysGlnHisIleLysLysGlnLeuSerArgPheAla 435

Qy  1441 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCGAAGCTCAGCGAAGGATGGTCATC 1500
          |||||||||||||||||:|||||||:|||||:|||||:|||||:|||:|||||:|||
Db   436 GlyAlaSerIleLysIleAlaProAlaGluGlyGluAlaGlnArgArgMetValIle 455

Qy  1501 ATCACCGGGCCACCGGAAGCCCAGTTCAACGGCGGCGACCGGATCTTTGGGAAACTGAAA 1560
          |||||:|||:||||||||||:|||||||||:|||:|:|:|||:|||||:|||:|||:|||
Db   456 IleThrGlyProProGluAlaGlnPheLysAspAlaGlyAspArgIleTyrGlyLysIleLys 475

Qy  1561 GGAGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGGTGGCAATATCAGATGTGCCC 1620
          |||||||||||||:||||:|:||||:|||||:|||:|||||||||:||||||:|||
Db   476 GlyGluAsnPheValSerProLysGluGluValLysLysGlyAlaHisIleArgValPro 495

Qy  1621 TCTTCCACAGTGCCAGAAGTCATCGTGCCTCGTGACACAGATTGATTTCCCGGCTGCCC 1680
          |:|||:||||||:||:|||||||||||::|||:|||||||||||||||:||||||:|||
Db   496 SerPheAlaAlaArgValAspIleLysGlyLysValThrValAsnGluLeuGlnAsn 515

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACACAAACGCCAGATGAAAAATGAGGAAGTG 1740
          |||||||:||||||||||||:|:|||||:|||||:|||||||||||||||||||||:|||
Db   516 LeuSerSerAlaGluValValValProArgAspArgGlnThrProAspGluAsnAspGlnVal 535
```

```
Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTGCTAGCCAGACGCACAGCGCAAGATCAGGGAA 1800
          :::||||||||  |||||||||||:::|||   |||   ||||||||||||:::|||
Db    536 ValValLysIleThrGlyHisPheTyrAlaCysPheTyrAlaCysGlnValAlaGlnArgGlnIleGlnGlu 555

Qy   1801 ATTGTACACAGGTGAAGCAGACAGGAGCAGAAA 1833
          :::||||||||||||||||:::||||:::
Db    556 IleLeuThrGlnValLysGlnHisGlnGln 566

RESULT 15
Q8IGK4
ID   Q8IGK4           PRELIMINARY;    PRT;    580 AA.
AC   Q8IGK4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RE72930p.
GN   IMP OR CG1691.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=y;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001739; AAN71494.1; -.
DR   FlyBase; FBgn0030235; Imp.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE 580 AA; 63407 MW; 86EA1F5D64523C02 CRC64;

Alignment Scores:
Pred. No.:                              580
Score:              3.09e-56   Matches:   231
Percent Similarity: 51.30%     Conservative: 104
Best Local Similarity: 35.38%  Mismatches: 187
Query Match:        14.27%     Indels:    131
DB:                            Gaps:      18

US-09-270-437D-6 (1-3412) x Q8IGK4 (1-580)
```

```
Qy     71 AsnProTyr------------------------------ProGlyMetProGlyProGlyArgGln 82
             :::                                    |||||
Db     71 AsnProTyr------------------------------ProGlyMetProGlyProGlyArgGln 82

Qy    643 ATTGATTCCGCGTCGCGGATCCTGTCCCACCCAGTTGTTGTGCCATCATCGGAAAG 702
          |||||||:::    |||||||||||||||||||:::|||
Db     83 AlaAspPhePheProLeuArgIleLeuValGlnSerGluMetValGlyAlaIleIleGlyArg 102

Qy    703 GAGGGCTTGACCATAAGAACTCACTAAGCACCAGTCCCGGGTAGATATCCATAGA 762
          ::: ||||||||||  ||||||||||||:::||||||
Db    103 GlnGlySerThrIleArgThrIleThrGlnSerArgAlaArgValAspValHisArg 122

Qy    763 AAAGAGAACTCTGAGCTGCGAGAGAAGCCTGTCACCATCCATGCACCCCAGAGGGGACT 822
          ||||||||  ||||||:::   |||
Db    123 LysGluAsnValGlySerLeuGlyLysSerIleThrIleTyrGlyAsnProGluAsnCys 142

Qy    823 TCTGAAGCATGCCCATGATTCTTGAAATCATGCAGAAATGCAGAGGCAGATGAGACCAAACTA 882
          ::: ||||||||||:::||||||
Db    143 ThrAsnAlaCysLysArgIleLeuGluValMetGlnGlnAlaIleSerArgThrAsnLys 162

Qy    883 GCCGAA---------------GAGATTCCTCTGAAAATCTTGGCACCAATGGCTTG 924
          ||||||
Db    163 GlyGluLeuSerProGluCysSerGluIleCysLeuLysIleLeuLeuAlaHisAsnAsnLeu 182

Qy    925 GTTGGAAGACTGATTGGAAGAAGAGGCAGAAAATTGAGAAAATTGAACATGAAACAGGG 984
          ::: ||||||||||||||:::|||
Db    183 IleGlyArgIleIleGlyLysGlyAsnThrIleLysGlyArgIleMetGlnAspThrAsp 202

Qy    985 ACCAAGATAACAATCCATCTTGCAGGATTTGAGCATATAACAACCCGGAAAGAACCATC 1044
          |||||||:::|||   ::: |||
Db    203 ThrLysIleThrValSerSerIleAsnSerPheAsnLeuGluArgIleIle 222

Qy   1045 ACTGTGAAGGGCCACAGTGAGGCCTGTGCCAGTGCGATAGGATTATGAGAAGCTG 1104
          ||||||||  ||||||:::   |||
Db    223 ThrValLysGlyLeuIleGluAsnMetSerArgAlaGluAsnGlnIleSerThrLysLeu 242

Qy   1105 CGTGAGGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCAATCTGATCCCAGG 1164
          ||:::|||||||||||:::||||   |||
Db    243 ArgGlnSerTyrGluAsnAspLeuAlaIleAlaProGlnSerLeuMetPheProGly 262

Qy   1165 TTGAACCTTCAGCGCACTT------------GGCATCTTTTCAACAGGACTG 1203
          |||||||||||||||   ::: |||
Db    263 LeuHisProMetAlaMetSerThrProGlyAsnGlyMetValPheAsnThrSerMet 282

Qy   1204 TCCGTGCTATCTCCACCAGCAGCAGGGCCCGCGGGAGCTCCCCGCTGCCCCTACCACCC 1263
          |||||||||||||||||||:::|||
Db    283 ------------------------------------------------Pro 283

Qy   1264 TTCACTACCCACTCCGGATACTTCTTCCAGCCTGTACCCCCATCACCAGTTGGCCCG-- 1320
          |||  ||||||||||::: |||
Db    284 PheProSerCysGlnSerPheAlaMetSerLysThrProAlaSerValValProProVal 303

Qy   1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAG 1380
          |||||||||||||:::|||
Db    304 PheProAsn-----------AspLeuGlnLeuThrPheTyrLeuTyrIleProAsnAsn 319

Qy   1381 GCTGTGGGGCCCATCATCGGGAAGAAGAGGGCACACATCAAACAGCTGCGAGATTCGCC 1440
          |||||||  ||||||||||:::|||
Db    320 AlaValGlyAlaIleIleGlyArgGlySerHisIleArgSerIleMetArgPheSer 339

Qy   1441 GGGAAGGCCTATCAAGATGCCCCT----------GACGTCAGCGAA 1488
          ||||||||||||||||||  :::|||
Db    340 AsnAlaSerLeuLysIleLeuAlaProLeuAspGlnIleThrGlu 359

Qy   1489 AGGATGGTCATCATCACCGGGCACCCAGTTCAAGGCCCAGGGACCGGATCTTT 1548
          |||:::||||||||||||   ::: |||
Db    360 ArgLysValThrIleValGlyThrProGluGlyInTrpLysAlaGlnTrpMetIlePhe 379

Qy   1549 GGGAAACTGAAAGAGGAGGAAAACTTCTTAACCCCAAAGAAGAAGTGGAAGCGCAT 1608
          |||||||||:::||||||
Db    380 GluLysMetArgGluGluGluAsnPheMetCysGlyThrAspSerValArgLeuThrValGlu 399

Qy   1609 ATCAGAGTGCCCTCTTCCACAGCTGGCCGGTGATTGGCAAAGGTGGCAAAGACCGTGAAC 1668
          ::: |||||||||||||:::
Db        
```

```
Db   400  LeuLeuValAlaSerSerGlnValGlyArgIleIleGlyLysGlyGlyGlnAsnValArg  419

Qy  1669  GAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCCGTGACCAA---ACGCCAGAT 1725
          ||||||||         :::  ::: ::::||||             ::::|||
Db   420  GluLeuGlnArgValThrGlySerValIleLysLeuProGluHisAlaLeuAlaProPro  439

Qy  1726  GAA-----AATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACT 1779
          :::|||||          |||||||    |||||||:::|||:::
Db   440  SerGlyGlyAspGluGluThrProValHisIleIleGlyLeuPheTyrSerValGlnSer  459

Qy  1780  GCCACAGCGCAAGATCAGGGAAATTGTA--------------------------- 1806
          ||||||||::::|||||||         :::::
Db   460  AlaGlnArgArgIleArgAlaMetMetLeuSerThrAsnProProIleThrLysLys  479

Qy  1807  ----------CAACAGGTGAAGCAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCC 1851
                    ::||||::::|||||||||||:::|||        ::||||:::
Db   480  GlnLysAlaAlaLysGluGluGlnGlnGlnGlnGlnSerLeu--AlaGlyAlaAlaS  499

Qy  1852  TCACAGCGCAGTGAGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCC 1911
          :::  :::     ||||::::||||:::        -----AlaL  516
Db   499  erSerGlySerGlnGlnGlnGlnProGlnSerProSerGlnGlnGln------

Qy  1912  TTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGA 1971
          ||||||       ::::||||        |||||
Db   516  euProProGlnLeuHisHisGlnProValSerSerAlaSerSerSerThrPro----  534

Qy  1972  CCATCTGAGGAATGAGAAGTCTGCCGAGGCGGCCCAGGGACTCTGCCGAGCCCTGAGAAC 2031
                                                   |||        :::
Db   535  ---------------------------------------ProAlaHisHisGlnG  540

Qy  2032  CCCAGGGGGCCGGAGGGGGCCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCG 2091
          |||        :::|||        :::|||
Db   540  lnGlnAlaSerThrAlaAlaThrSerHisGlnLeuGlnGlnGlnGlnProSerProPro-  559

Qy  2092  CCTCCCGCCCCCAGGGCTTCTGCAGGCTTCAGCC 2126
          |||||||         ||::::|||
Db   560  -------ProProGlyAsnAlaThrAlaAlaAla  568
```

Search completed: July 16, 2004, 10:44:26
Job time : 203.5 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: July 20, 2004, 21:16:56 ; Search time 8261 Seconds
(without alignments)
12333.829 Million cell updates/sec

Title: US-09-270-437D-6
Perfect score: 3412
Sequence: 1 ggcagcggaggaggcgagga........aaccttgaaaatgttattt 3412
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2751328289 seqs, 1431090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssI:*
29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2163	63.4	3592	11	AK049196	AK049196 Mus muscu
2	2006.2	58.8	3262	11	AK077404	AK077404 Mus muscu
3	1862.8	54.6	2964	11	AK044984	AK044984 Mus muscu
4	842.2	24.7	874	13	BQ691878	BQ691878 AGENCOURT
5	841	24.6	905	12	BM453327	BM453327 AGENCOURT
6	830	24.3	851	13	BQ221568	BQ221568 AGENCOURT
7	829.8	24.3	999	12	BM561242	BM561242 AGENCOURT
8	809	23.7	832	13	BX644668	BX644668 DKFZp781G
9	803.4	23.5	1089	12	BM806132	BM806132 AGENCOURT
10	799.2	23.4	836	13	BQ212500	BQ212500 AGENCOURT
11	759.8	22.3	894	13	BU156245	BU156245 AGENCOURT
12	752.6	22.1	781	12	BG575889	BG575889 602598315
13	751.4	22.0	940	14	CD243351	CD243351 AGENCOURT
14	749.8	22.0	824	14	CF593505	CF593505 AGENCOURT
15	746	21.9	889	12	BG748346	BG748346 602705902
16	718.8	21.1	762	13	BU601634	BU601634 AGENCOURT
17	708.6	20.8	4100	11	AK088465	AK088465 Mus muscu
18	707	20.7	2202	11	AK011689	AK011689 Mus muscu
19	698.2	20.5	2780	11	AK013940	AK013940 Mus muscu
20	695	20.4	2718	13	AK044850	AK044850 Mus muscu
21	691.8	20.3	939	13	BQ961314	BQ961314 UI-CF-DU1
22	689	20.2	724	12	BM977219	BM977219 UI-CF-DU1
23	686.4	20.1	891	13	BU152527	BU152527 AGENCOURT
24	677.8	19.9	918	13	BQ212353	BQ212353 AGENCOURT
25	652.6	19.1	815	13	BI688088	BI688088 603314678
26	634	18.6	647	13	BU674692	BU674692 UI-CF-DU0
27	631.6	18.5	729	10	BE278075	BE278075 601179604
28	625.4	18.3	859	13	BQ890899	BQ890899 AGENCOURT
29	606.2	17.8	759	10	BE622021	BE622021 601440673
30	605.4	17.7	1024	9	AL549055	AL549055 AL549055
31	599.6	17.6	821	14	CA511280	CA511280 UI-R-FJ0-
32	595	17.4	628	12	BI862502	BI862502 603390625
33	594	17.4	607	12	BM724394	BM724394 UI-E-E01-
34	591.8	17.3	759	14	CA320888	CA320888 UI-M-FW0-
35	587	17.2	612	13	BU679082	BU679082 UI-CF-DU1
36	584	17.1	584	13	BX642777	BX642777 DKFZp781I
37 c	571.4	16.7	616	14	CA430674	CA430674 UI-H-FL1-
38	565	16.6	638	12	BM843467	BM843467 K-EST0121
39	561.2	16.4	699	13	BU705477	BU705477 UI-M-FO0-
40 c	553.8	16.2	1078	12	BM462051	BM462051 AGENCOURT
41	548	16.1	594	12	BI858380	BI858380 603383965
42	547.4	16.0	1212	13	BU902212	BU902212 AGENCOURT
43	546.8	16.0	668	13	BY734118	BY734118
44	542.6	15.9	845	13	BU155790	BU155790 AGENCOURT
45	541.6	15.9	707	13	BY720350	BY720350 BY720350

ALIGNMENTS

RESULT 1
AK049196
LOCUS   AK049196   3592 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
            clone:C330012H03 product:HEPATOCELLULAR CARCINOMA AUTOANTIGEN
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK049196
VERSION   AK049196.1 GI:26093366
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE 99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE 20499374
   PUBMED   11042159

REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,A., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3592)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..3592
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C330012H03"
                     /db_xref="MGI:2416441"
                     /db_xref="taxon:10090"
                     /clone="C330012H03"
                     /cell_type="ES cells"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     misc_feature   34..1811
                     /note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
                     sapiens] (SPTR|Q9Y6M1, evidence: FASTY, 88.3%ID,
                     100%length, match=1775)
                     putative"
ORIGIN

Query Match   63.4%;   Score 2163;   DB 11;   Length 3592;
Best Local Similarity   82.9%;   Pred. No. 0;

Matches 2831;   Conservative   0;   Mismatches 486;   Indels 97;   Gaps 28;

```
Qy        38 GGGGGAGCGCGGGCTCTCGGGAAGAGACGGATGAGAAGCAAGCTTTACATCGGGAACC   97
             ||||||||||||||||||||||  ||||||||||||||||||||||||  |||||||||
Db         3 GGGGGAGCGCGGGCTCTC-GGGAAGACGCGGATGATGAGCAAGCTGTACATTGGGAACC   61

Qy        98 TGAGCCCCGGCCGTCACCGGCCACCGACCTCCGGCCAGCTCTTTGGGACAGGAAGCTGCCCC   157
             |||||||||  |||||||||||||||||||| |||||||||||||||||||||||||||||
Db        62 TGAGCCCCGGCCGTCACCGGCCACCGACCTCCGGCCAGCTCTTCGGGACAGGAAGCTGCCCC   121

Qy       158 TCGGCGGGACAGGTCCTGCTGAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACT   217
             ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       122 TGGCGGGACAGGTCCTACTCAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACT   181

Qy       218 GGGCCATCCGCCGCCATCGAGACCCTCGGGTAAAGTGGGATTGCATGGGAAAAATCATGG   277
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       182 GGGCCATCCGCCGCCATCGAGACCCTCGGGTAAAGTGGGATTGCATGGGAAAAATCATGG   241

Qy       278 AAGTTGATTACTTCAGTCTCTAAAAAGCTAAGGAGCAGGAGAAAATTCAGATTCGAAACATCG   337
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       242 AAGTTGACTACTTCAGTCTCTAAAAAGCTAAGGAGCAGGAGAAGAATCCAGATTCGGAACATCC   301

Qy       338 CTCCTCACCTGCAGTGGGAGGTGTTGGATGGGACTTTGGCTCAATATGGACAGTGGAGA   397
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       302 CGCCTCACCTGCAGTGGGAGGTGTTGGATGGGCTGTTGGCTGAATATGGACAGTGGAGA   361

Qy       398 ATGTGGAACAAGTCAACACAGACAGAAACCGCCGTTGTCAAGTCACATATGCAACAA   457
             ||||||||||||||||||||||||||||||||||  |||||||||||||  |||||
Db       362 ACGTGGAGCAAGTCAACACAGATACAGAAACTGCCGTTGTCAAGTCACCTATATGACAA   421

Qy       458 GAGAGAGAGCAAAAATAGCCATGGACGAAGAGCTAAGCGGGCATCAGTTGAGAACTACTCCT   517
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422 GAGAGAGGCAAAGCTAGCTATTGAGAAGCTCAGTGGGCATCAGTTTGAGACTACTCCT   481

Qy       518 TCAAGATTTCTACATCCCGATGAAGAGAGTGAGTCCCCTTCGCCCCTCAGCGAGCCC   577
             ||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       482 TCAAGATTTCCTACATCCCCGATGAAGAGGTGAGCTCTTCCTTCACCCCCTCATCG-----   536

Qy       578 AGCGTGGGAGCACCACCTCTTCCCGGAGCCAGGCCAGGCCACGCCCTGGGGCCACTTCTCAGGCCA   637
             ||||||||||||||                                  ||||||||||||||||||
Db       537 -------TGCCCGGGAACAAGGCCACGGCCCACGGCCCGGGAGCTCTTCTCTCCAGGCCA   580

Qy       638 GACAGATTGATTTCCCGCTGCGGATCCTGGTTCCCCACCCAGTTTGGTGTCCATCATCG   697
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 GACAGATTGATTTCCCGCTGCGGATCCTGGTTCCCCACCCAGTTTGGTGTCCATCATCG   640

Qy       698 GAAAGAGGAGGCTTGACCATAAGAACATCACTAAGCAGAACATCCCGGGTAGATATCG   757
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       641 GAAAGGAGGAGGCTTGACCATAAGAACATCACTAAGCAGAACATCCCGGGTAGACATCC   700

Qy       758 ATAGAAAAGAGAACTCGGAGCTGCAGAGAGGCAGATGAGGCACCCCAGAGG   817
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       701 ACAGAAAGGAGAACTCTGGGGCTGCAGAGAGGCAGATGAGGCACCCCAGAAG   760

Qy       818 GGACTTCTGAAGCATGCGCCATGATGATTCTTGAAATCATGCAGAAATCATGCAGAAGATGAGACCA   877
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 GGACATCTGAAGCATGCGCCATGATGATTCTTGAGATTCTTGAAAATGCAAAAGAAGCTGATGAGACCA   820

Qy       878 AACTAGCCGAAGAGATTCCTCTGGACACCAATGGCTTGGTTGGAAGACTGA   937
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 AACTGGCTGAGGAGGTTCCTCTGGAAATCCTGGCCCACAATGGCTTCGTTGGAAGACTGA   880

Qy       938 TTGGAAAGAGGAGGCAGAAATTGAAGAAATTGAACATAGAGAAAGAGATAACAA   997
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 TTGGCAAAGAGGAGGCAGAAACCTGAAGAAAATAGAGAAAGAGAAATAACCA   940

Qy       998 TCTCATCTTTGCAGATTGACATATACAACCCGGAAGAACCATCACTGTGAAGGGCA   1057
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 TCTCATCCTTGCAGATTGAGCATTTATAACCCCGAGAGAACCATCACCGTGAGGGGCA   1000

Qy      1058 CAGTTGAGGCCTGTGCCAGTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCCTGAGGCCTTTG   1117
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 CCATTGAAGCCTGTGCCAATGCTGAGATTATGAAGAAGAGCTCCGAGAGGCCTTTG   1060
```

```
Qy   1118 AAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCG 1177
          |||||||| |||||||||  ||||||||||||||||||||||||||||| ||||||||
Db   1061 AGAAGCAGACATGCTGGCCGTTAACCAACAAGCCAATCTGATCCCAGGGCTAAACCTCAGTG 1120

Qy   1178 CACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCACCCACTCCGATACTTCTCCAGCCTGT 1237
          |||||||||||||| |||  |||||| ||||| ||||||||| ||||  ||| ||||||||||||| | |
Db   1121 CACTTGGCATCTTTCGACTCTTCGGACTGTCGTGCTGTCTCCCTCCCA-CAGCCAGGGCCCGTGGAG 1179

Qy   1238 CTCCCCCGCTGCCCCCTACCACCCCTTCACTACCCACTCACTCTTATCGATCCTTCTCCAGCCTGT 1297
          ||     |||| ||| |||||| ||  |||    |||||| ||||  |||||| |||| ||||||||
Db   1180 TTCCCCCCAGTCCTCCCTATCACCCCTTGCTACCCCACTCACTCCTCCGGATACTTCTCCAGTCTGT 1239

Qy   1298 ACCCCATCACCAGTTGGCCCGTTCCCGCATCATCATCTCTATCCAGCAGCAGGAGATTG 1357
          |||||||||||  |||| |  |||| ||||||||||| || |||||||||||||||||||
Db   1240 ACCCTCATCACCAT-TTCGGCCGTTCCCACCATCACTCC-ACCAGCAGCAGGAGGACTG 1299

Qy   1358 TGAATCTCTTCATCCCAACCCAGGCTGTGGCGCCCATCATCGGGGAAGAGGGGCACACA 1417
          ||  ||||||||  | ||||||| || |||||||||||||| || ||||||||||| ||
Db   1300 TAAGTCTCTTCATCCCAACCCAGGCTGTATCATCGGGTATATCGGGAAGAGGGGCACACA 1359

Qy   1418 TCAAACAGCTGGCCGAGATTGCCCGGAGCCTCATCATCAAGATTGCCCTCGCGGAGGCCCAG 1477
          |||||||||||||||||||||||||||||| ||||||||||||||| |||| | ||||||||
Db   1360 TCAAACAGCTGGCCGATCGATTGCCCGGAGCCTCATCAAGATTGCTCCAGCAGAAGGTCCAG 1419

Qy   1478 ACGTCAGCGAAAGGATGGTCATCATCACCGGGCCACCGGAAGCCAGTTCAAGGCCCAGG 1537
          |||||||| |||||||||||||| ||||||||||||||||||||||||||| ||||||||
Db   1420 ATGTCAGTGAGAGGATGGTCATCATCACTGGTCCTCCTGAAGCCCAGTTTAAGGCTCAGG 1479

Qy   1538 GACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGTGAGTGAAGC 1597
          |||||| || |||||||||||||||||||||| ||| ||| ||||||||| ||| ||||||
Db   1480 GACGGATCTTTGGGAAACTGAAAGAGGAAAAACTCTTTAATCCCAAAGAAGTGAGTGAAGC 1539

Qy   1598 TGGAAGCGCATATCAGAGTGCCCCTTCCACAGCTGGGCGGTGATTGGCAAGGTGGCA 1657
          |||||||||||| ||||||| |||||||||||||||| || |||| |||||||||||
Db   1540 TGGAAGCGCCACATCCGAGTCCCCATCGCTGGGCGGTGATTGGCAAGGGGCGGGA 1599

Qy   1658 AGACCGTGAACGAACTGCAGAACTTAACCAGTGCGAGAACTCATCGTGCCTCGTGACCAAA 1717
          |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db   1600 AAACCGTGAACGAACTGCAGAACTTGACCAGTGCGAGAACTCATCGTGCCTCGTGACCAAA 1659

Qy   1718 CGCCAGATGAAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGA 1777
          ||||||||| ||||||||||||||| |||| |||||||||||| ||||||||||||||||||
Db   1660 CGCCAGAATGAGGAAGTGATCGTCAGAATTATCGGGCATTTTTTTGCTAGCCAGA 1719

Qy   1778 CTGCACAGCCGCAAGATCAGGGAAATTGTACAACAGTGAAGCAGCAGAATACC 1837
          ||||||||||||||||||||||||||| ||| |||||||||||||||| |||||
Db   1720 CTGCACAACGGCAAGATCAGGGAAATTGTTACGAGCAGTGAAGCAGCAGAGATACC 1779

Qy   1838 CTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACACG 1897
          ||||||||||||||||| ||||||||||||||||||| ||||  ||||||||||||||
Db   1780 CTCAGGGAGTCGCCACAGCGCAGCAAGTGAGGCTGAGGCTC-CACCAGCAAGCAACC 1837

Qy   1898 GATGAATGTAGCCCTTCCAACACCTGACCAGAATGAGACCAGCCAGCCAGCAGATCGG 1957
          ||||||||||||||| | |||||||||||  ||| ||||||||||||||||||||||||
Db   1838 GATGAATGTAGCCCTCCCCCAACACCTGACAG-ATGAGACCAAAC--AGCCAGCAGATCGG 1893

Qy   1958 GAGCAAACCAAAGACCAT-CTGAGAATGAGAAGTCTGCGGAGGGGGCAGAAATACC 2016
          |||||||||| ||||  | |||||||||| ||||||||||||||||||||| ||||
Db   1894 GAGCAAACCAAAGAGCAGCATCCCAACAGTTCGCCGGGGGCTCTGCGAGAGCCAGCAGAGATCC 1950

Qy   2017 CGAGGCCCTGAGAACCCAGGGGCCGAGGGGCCGGAAGGTCAGCCAGGTTTGCCAG 2076
          ||||||| ||||||||||||||  |||||||||||||  ||||||||||||||||
Db   1951 --AGACCTCTACACATCCTGGATCCCTGGGATCCAGGGGCACCAGGGAAG--GCCCAGGTTGTCCAG 2005

Qy   2077 AACCACCG--AGCCCGCCTTCTGCCATCCAC--CAGCATCCAC 2132
          |||||||   ||||||||||||||||  ||   |||||||  |
Db   2006 AAACACCGCTTGGCCTGCCTTCTGCCCCTGGCTTCTGCAGGCATAACAGCCATCCAC 2065

Qy   2133 TTCACCATCCACTCGGATCTCTCCTGAACTCCCTTTTAGTTGAACTA 2192
          ||||||||||| ||||| || ||||||||||||||| ||||||||||
Db   2066 TGCCCATCCAACTCAGAT-TCTCCCAGTTCCCAGGACCGCTATCCCTTTCGGTTGTTGAACTA 2124
```

```
Qy   2193 ACATAGGTGAACGTGTTCAAAGCCAAGCCAAGATGCACACC----TTTTCTGTGTGGCAAA 2248
          ||||||||||||| ||||||||||| |||||||  |||||||    |||| || ||| |||
Db   2125 ACATAGGTGAACATGCTCAAAGCAAAGCCAAAATTCCTAGCCCGTTTCTTGTGTGTGGAAAG 2184

Qy   2249 TCGTCTCGTACATGTGTGTACATATTAGAAGGGA-GATGTTAAGACATATGGCCTGT 2307
          ||| ||  |||| ||| |||| |||||||| ||| |||||| ||| |||| |||||||
Db   2185 TTGTCTCGTTACATGTATGTACATAAATCAGAAGGGGTTAAGAAATGTTAAGAATGTGGCCTGT 2244

Qy   2308 GGGTTACACAGGGTGCCTGCAGCGTGTAATATATTTTGAGAAATAATATATCAAATAACTCA 2367
          ||||||||||||||||||||||||||| |||||| |||||| || ||||| ||||||| | 
Db   2245 GGGTTACACAGGGTGCCTGCAGCGTGTAATATTTTTAGAAATAATATATCAAATAACTCC 2304

Qy   2368 ACTAACTCCAATTTTAATCAATTATTAATTTTTTTTCTTTTAAAGAAGCAGGCT 2427
          ||||| |||||||||| ||||| |||||| ||| |||||  ||||||||||||||
Db   2305 ACTAACAACAATTTTTACTCAGTGATTAATTTTGTTTTTCTTTTTGAAGAGAAAGCAGGCT 2364

Qy   2428 TTTCTAGACTTTAAAGAATAAAGTCTTTGG-GAGGCTCACGTGTAGAGAGAGCTTTG 2486
          |||| ||||||||| |||||||||||| ||| |||||||||||| |||||||| || 
Db   2365 TTTCTAGACTTCCAAGAATAAAGTATTGGGTGAGGTCCACTGGTATAGAAAGGAACTTG 2424

Qy   2497 AGGC-CACCCGCACAAAATTCACCCGAGGAAATCTCGTGGAAGGACACTCACGGCAG 2545
          ||||  |||||| ||||| ||||||| |||||||| || | |||||| |||||| |||
Db   2425 AGCCACCCCACACAAAAATTCACTGGAGGGAAATCTTGTC-AAGGACACTTAATGGCAG 2483

Qy   2546 TTCTGGATCACCTGTGATGTCAACAGAGGGGATACCGTCTCCTTGAAGAGAAACTCTG 2605
          || |  |||||||| ||||| |||||||||||| ||||||||||||||||||| |||||
Db   2484 TTGTGGGTCACCTGTGTCTGTCAACAGAAGGGATAACCGTCCCTGAGGAGGGGCCCTAGGC 2543

Qy   2606 TCACTCCTCATGCCTGTCTAGCTCAGCAGCTCATACACCCCATTCTCT- 2645
          |||||| || | ||||| ||| || | |||||| |||||||||||||||
Db   2544 CCTGGTCACCTGT-TCTCCCCACACCCATTGCTCAGTTTCACAGGTTTTTAAACTGATG 2603

Qy   2646 TTGCTTCACAGGTTTAAACTGTTGTTTTTTGCATACTGCTATATTCTCTGTCTCTCTC 2705
          |||  ||||  ||| |||||||| || || |||| ||||||||| ||||||| ||||
Db   2604 ATTTTGGTTTGTTTGTTTTGTTTTTTGCATACTGCTATATAGTTCTGTCTCTCTC-- 2662

Qy   2706 TGTTTATCTCCCCTCCCCTCCCCTCCCCTTCTTCTCCATTCTTTTTGAATTTCC 2765
          |||| ||||| ||| || | ||  |||| |||||||| ||||  ||||||  |||
Db   2663 -----TTATCTCCCTCCTCTGTCTGTCTTCCTCCGTTCTTTTACCCAAAATTCTTTGAGTCCC 2717

Qy   2766 TCATCCCTCCATCTCAATCCGTATCTACGCACCC---CCCCCCCCCAGGCAAAGCAGT 2822
          ||| | ||| |||| |||| |||| ||||| |||   ||||||||||||||||||||||
Db   2718 TTGTTCTTTTTCTGCTCCCGTATCTATGCACCCCCTCCCCTCCCCCCAGGCAAAGCAGT 2777

Qy   2823 GCTCTGAGTATCATCACCATCACCACAAAGGAACACAAAGCAACACAACCAGCCTCAACT 2882
          |||||||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||
Db   2778 GCTCTGAATCATCATCACCATCACCACCACAAAGCAACACACAAGTGCAACCAGCCTCAGCT 2837

Qy   2883 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGCGTCGGCGGTACTTGTCCTAGGAGAG 2942
          || | |||||||| |||||||||||||| ||||| |||||| |||||||||||||||||||
Db   2838 TACACTCGGTTACTCAAGCGAACAAGAGTCGGCGGTACTTGTCCTAGC-ATTCGGAAGAG 2896

Qy   2943 GAAAACAGGAACCCACCACCAACCAATCAACCAAACAAAAATTCCACAATGAA 3002
          |||||||||||| ||| | |  |||||| |||||||| | | || ||| |||||
Db   2897 GAAAGCAGGATCCCAGAGCCCAGCAAATAAACCAGCCAAACAAAG-AAAAATCCCCACAAAGA 2955

Qy   3003 AGAATGTATTTTGCTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3062
          ||||||||| |||| | ||| ||| |||||||||| |||| ||||||||||||||| ||||
Db   2956 AGAATGGA----CTCTCTCCATCTTGGTATACAAGGCCATCAACATTTAGTGAAATCT 3011

Qy   3063 TCCTTTCTTTAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCA 3122
          |||| || || ||||  ||||||| |||||| ||||| | ||||||||||||||||||
Db   3012 TTCTTCTTTTAAAAATTTTAAAGTGCAGGAAAAATATGCAATTTTATCGGAGGTTGTTGGCCCA 3071

Qy   3123 GGGCGGTAAATCACACAGATTTTTTAACAGAAAAACACACAGA---AGAAGTACCTCA 3179
          |||| |||||||||||||||| |||| |||| | |||||| |||    ||||||||||||
Db   3072 GGGCGGTAAATTACAGATTTTTTTTAAATGAGAAAAACACACCATCCACTACCTCA 3131

Qy   3180 GGGTG-TTTTTACCTCAGCACCTTGTCTCTTTTAGTTGAACTA 3238
          |||| |||||| |||||||||| |||||||||| |||||||||
Db   3132 GGTGTTTTTTACCTCAGCACCTTGTTTCCCTTAGAGATTTGTAAAACTGAT 3191

Qy   3239 AGTTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAA 3298
```

```
Db   3192  |||||||||  ||||||||  |||||||||||||||||  ||||||||||||||  ||  |||
           AGTTGGAGCA-TTTTTTATTTTTTAATAAAATGAGTT-GAAAAAAAAAAAAGATGAG 3249

Qy   3299  CTGCCAGCCTGGAGAGGTGCAGTGCAGTGTGCAACAGCTGTTCTGAATTGTCTTCCG 3358
           |||||||||||  |||||||  ||||||||  |||||  |||  ||||||||||
Db   3250  CTGCCAGCCCCAAGCAGCGGCGCAGCCCAGGTGTGCCAGCTCTTC-----TGTCTTCCG 3304

Qy   3359  CTAGCCAAGAACCNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
           |||||||| ||  ||||||||||||||||  |||||| ||  || |||||
Db   3305  CTAGCCAAGAGCCTACGTGCCTTCTTCTGTGGACAAAACCTAAAAATGTTTATTTT 3358
```

RESULT 2
AK077404

LOCUS   AK077404   3262 bp   mRNA   linear   HTC 18-SEP-2003
DEFINITION Mus musculus 6 days neonate head cDNA, RIKEN full-length enriched
            library, clone:5430439D08 product:HEPATOCELLULAR CARCINOMA
            AUTOANTIGEN homolog [Homo sapiens], full insert sequence.
ACCESSION   AK077404
VERSION   AK077404.1 GI:26097330
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3262)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,J., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numata,K., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
```

```
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3262
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5430439D08"
                     /db_xref="MGI:2393565"
                     /db_xref="taxon:10090"
                     /clone="5430439D08"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="6 days neonate"
     misc_feature    16..1796
                     /note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
                     sapiens] (SPTR|Q9Y6M1, evidence: FASTY, 88.3%ID,
                     100%length, match=1775)
                     putative"
     polyA_signal    3249..3254
                     /note="putative"
     polyA_site      3262
                     /note="putative"
ORIGIN
Query Match        58.8%;   Score 2006.2;   DB 11;   Length 3262;
Best Local Similarity 81.5%;   Pred. No. 0;
Matches 2693;   Conservative   0;   Mismatches   483;   Indels 127;   Gaps   27;

Qy    58  GGGAAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGGTCACCGCC 117
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4  GGGAAGAGACGGGATGATGAACAAGCTGTCATTGGGAACCTGAGCCCCGGTCACCGCC 63

Qy   118  GACGACCTCCGGCAGTCTTTGGGACACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTG 177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64  GACGACCTCCGGCAGTCTTCGGGGACAGCAAGCTGCCCCTGGCGGGACAGGTCCTCACTC 123

Qy   178  AAGTCCGGCTACGCCTTCGTGGACTACCCGACCAGAGCTGGGCCATGGGCCATCGGAG 237
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124  AAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCCATCGAG 183

Qy   238  ACCCTCGGGTAAAGTGGAATTGCATGGGAAGAAATCATGGAAGTTGATTACTACTGTCT 297
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184  ACCCTCGGGTAAAGTGGAATTGCATGGGAAGAAATCATGGAAGTTGACTACTCAGTCTCT 243

Qy   298  AAAAAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCTCAGTGGGGAG 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244  AAAAAGCTAAGGAGCAGGAGAATCCAGATTCGGAACATCCGCGGCCTCACCTCAGTGGGGAG 303

Qy   358  GTGTTGGATGGACTT-TTGGCTCAATATGGGAACAGTGGAACAAGTCAACAC 416
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   304  GTGTTGGATGGGCTTGTTGTTGGCTGAATATGGGACAGTGGAGCAAGTCAACAC 363

Qy   417  AGACACAGAAACCGCGTTGTCAACGTCACATATGCACATTGCACAAGAAGAACAAAATAGC 476
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364  AGATACAGAAACTGCCGTTGTCAACGTCACCTATATGACCAAGAGAAGAAGAGCAAGCTAGC 423
```

```
Qy    1556 TGAAAGAGGAAAACTTCTTTACCCCAAAGAGAAGTGAAGCTGGAAGCGCATATCAGAG 1615
           |||||||||||||||| ||||||||  ||||||||||||||||||||||||||||||||
Db    1483 TGAAGGAAGAAAACTTCTTTAATCCCAAAGAGAAGAAGTGAAGCTGGAGGCCCACATCCGAG 1542

Qy    1616 TGCCCCTTCACAGTGGCTGGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGC 1675
           ||||||||||||||||| ||| | || |||||||  ||| |||||||||||||||||||||
Db    1543 TCCCCTGTCGACCGCGTGGCCCGGGTGATTGGCAAGGTGGCAGGGCGGGAAAACCGTGAACGAGCTGC 1602

Qy    1676 AGAACTTAACCAGTGACGAAGTCATCGTGCCTCGTGCCCAAACGCCAGATGAAAATGAGG 1735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1603 AGAACTTGACAGTGACGAAGTTATCGTGCCTCGTGACCAAACGCCAGAGAATGAGG 1662

Qy    1736 AAGTGATCGTCAGAATTATCGGGCACTTCTTTCTTGCTAGCAGCAGCGCAAGATCA 1795
           ||||||||||||||||||||||||||||  |||||||||||||||||||| ||||||||
Db    1663 AAGTGATCGTCAGAATTATCGGGCATTTTTTGCTAGCAGCAGCGCAACGGCAAGATCA 1722

Qy    1796 GGGAAATTGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1723 GGGAAATTGTACAGCAGTGAAGCAGCAGCAGGAGCAGAGATACCCTCAGGGAGTCGCCCCAC 1782

Qy    1856 AGGCCAGCAAGTGAGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCC 1915
           ||||||||||||||| |||||||||    ||||||||||||||||||||||| ||||||
Db    1783 AGGCCAGCAAGTGAGGTCCCACAG--CACCAGCAAGCAACCGGATCGGAGCAAACCAAAGACCAT 1840

Qy    1916 AACACCTGACAGAATGAGACCAAAACGCCAGCCAGCAGATGCGGAGGACTCTGCGGGAACCCC 1975
           |||||||||||||  ||||||||||  |||||| |||||||| ||||||||||||||||||
Db    1841 AACACCTGACAG-ATGAGACCAAAC--AGCCAGCAGATCGGGAGCAAACCAAAGAGCAT 1896

Qy    1976 -CTGAGGAATGAGAGTCTGCGGCGGCGGGAAGGTCAGCCAGTTTGCCAGAACACCACCG--AGCCCGC 2034
            |||||||||||||||||||||||||||||||||||||||||||      |||||||||  |||||||
Db    1897 CCGGAGGAGTGCCCAGTCTGCAGGAGGCCACAGAGGCAGGGCCTTCAGGTTGTCCAGAACACCCGCTTGGCCTGC 1952

Qy    2035 AGGGGCCCCAGGAGCTTCTGCGAGGCTT--CAGCCATCCATCCACTCGAT 2092
           |||||||||||||||| |||||||||||  ||||||||||||||||||||
Db    1953 TGGGATTCCAGGAGCTTCTGCGAGGCATAACAGCCATCCATCCCAACTCAGAT 2009

Qy    2093 CTCCCGCCCCCAGGGCTTCTGCGAGGCACTGTGAGGGCACAGCCTGTG--CCA 2150
           |||||||||||||||||||||||||||| ||||||||||||||||||||  |||
Db    2010 CCCCAGCTTCCCCCTGGCTTCTGCGAGGCATAACAGCCATCGCCCATCCCAACTCAGAT 2069

Qy    2151 CTCTCCTGAACTCCCACGACGCTATCCCCTTTAGTTGAACTAACATAGGTGAA--CGTGT 2208
           | ||||||||||||||||||||||||||||||||||||||||||| ||||||  |||||
Db    2070 -TCTCCTCAGTTCCCAGGACGCTATCCCCTTTCCGTGTGACTAACATAGGTGAACACATGCT 2128

Qy    2209 TCAAAGCCAAGCAAAATGCACACCCTTTTC-----TGTGGCAAATCGTCTCTGTACATG 2263
           ||||||||||||||||||||||||||||||     ||||||||||||||||||||||||
Db    2129 CAAAGGCCAAGGAAAAATTCCTAGCCGTTCTTTGTTGTGGAAGTTGTCTCTGTACATG 2188

Qy    2264 TGTGTACATATTAGAAAGGGAA-GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTG 2322
           |||||||||||||||||||||| ||||||||||||||||| |||||||||||||||||||
Db    2189 TATGTACATATCAGAAGGGGAAGGATGTTAAGAAATGTGCCTGTGGGTTACACAGGGTG 2248

Qy    2323 CCTGCAGCGGTAATATATTTAGAATAAAGAAGCAGCGTTTGAGGCACC 2382
           |||||||||||||||||||||||||||||||||| |||||||||||||
Db    2249 CCTGCAGCGGTTAATATATTTAGAAGAACAACAAACAGCGTTTGAGGCACC 2308

Qy    2383 TAATCAATTATTAATTTTTTTTCTTTTAAAGAAAGCAGAGTTTAAA 2442
           ||||||||||||| || ||||||||||||||||||||||||||||
Db    2309 TACTCAGTGATTTAATTTTGTTTTCTTTTTTGAAGAGAAAGGAACTTTGAGCCCACC 2368

Qy    2443 GAATAAAGTCTTTGGGAGGTCTCACGGT------GTAGAGGAGGCTTTGAGGCCCACC 2494
           |||||||||||||||||||||||||||      |||||||||||||||||||||||
Db    2369 GAATAAAGATTGGGTGAGGGTCCCACTTGGTATTAGAAAGGGAACTTTGAAGCCCACC 2428

Qy    2495 CGCACAAAA--------------------------TTCACCCAGAGGGAAATCTCGTCGGAA 2530
           ||||||||                          |||||||||||||||||||||||||
Db    2429 CACACAAAAAAAAAATTTCACTTGGAGGGAAAATTTTTTCAAGAACAGGGAATGCCC 2488

Qy    2531 GGACACTCACGGCAGTTCTGGATCACCTGTATGTCACAGAGGGATACCGTCTCCTT 2590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2489 AGTTTGTTGGGTTCCACCCTGTTTCTTCGTCCAAGCCAGAAAGGGAAAAACCCTTCCCT 2548

Qy    2591 GAAGAGGAAACTCTGTCACTCCTGTCTGTCAGTCTCAGTCATACACCCATTTCTTCTCTTTGCT 2650
```

TITLE
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20493374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2964)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagami,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal
Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
Kanagawa Prefecture, Japan) whose assistance we gratefully
acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1..2964
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:B130017M24"
/db_xref="MGI:2410341"
/db_xref="taxon:10090"
/clone="B130017M24"
/tissue_type="parthenogenote"

Db   2549 TAAGGGAAGGGGCCCAGGCCCTTGTCACCCGTTTCTCCCACACCCATTGCTCAGTT 2608
                  |||| | | | |||||||||||||||||||||||||||| |||| ||||
Qy   2651 TCACAGGTTTTAAAC------------------TGGTTTTTGCATACTGCT 2684
                  |||||||||| | ||||||| || |||
Db   2609 TCACAGGTTTTTAAATTAATAATTTTGGTTGTTTGTTTTTTGCATACTGCT 2668

Qy   2685 ATATATTCTCTGCTCTCTGTTTATCTCCCCGCCCTCCCCCTTCTTCTCCA 2744
                  ||||| ||||||||| |||||||||||||||||||||||||| |||||
Db   2669 ATATAGT--TCTGTCTCTCTCTATCTCCCTCCTAACCTTTCCTCGTCTTCTTTACCC 2726

Qy   2745 TCTCCATTCTTTTGAATTCC---TCATCCTCCCATTCCAATCCCGTATCTACCGCACC-- 2799
                  |||||||||||||||||||| |||||||||||||||||||||||||||||
Db   2727 CAGAAATTCTTTTTTGAGTCCCCTTTTCGTTCTTTTTCTGCTCCCCGTATCTATGCACCCC 2786

Qy   2800 -CCCCCCCCCCAGGCAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGC 2858
                   |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db   2787 TCCCCTCCCCCAGGCAAAGCAGTGCTCTGAAATATCATACCACACAAGGAACAAATGC 2846

Qy   2859 GAAACACACAAAACCAGCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGT 2918
                  ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   2847 GAACCACCAGACCAGCCCTCAGCTTACACTCGGTTACTCAAGCGAACAAGAGTCGGCGGT 2906

Qy   2919 ACTTGTCCTAGCGTTTGAAGAGAGAAAACAGGAAACCACCAACCAATCAACCAA 2978
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2907 ACTTGTCCTAGC-ATTCGGAAGAGAGGAAAGCAGGATCCCAGCAAATAAACCAGCCAGCCAA 2965

Qy   2979 ACAAAGAAAAATTCCACATGAAGAAATGTATTTTGTCTTTTGCATTTTGGTGTATAA 3038
                  |||||| ||||||||||||||| ||| |||||||||||||||||||||
Db   2966 ACAAAG-AAAAATCCCACAAAGAAAGAATGGA----CTCTCCTCCATCTTGGTATACAA 3020

Qy   3039 GCCATCAATATTCAGCAAAATGATTCCTTCTTTAAAAAAAAAATGTGGAGGAAAGTAG 3098
                  ||||||| |||| |||||| || |||||||| ||||| |||
Db   3021 GCCATCAACATTTTAGTGAAATTCTTTCTTCTTTCTTAAAAATTTAAAGTGCAGGAAAATAGC 3080

Qy   3099 AAATTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAA 3158
                  |||| ||||||||||||||||||||| |||||||||||||||||||||||
Db   3081 AATTTATACGGAGGTTGTTGCCCAGGGCGTTAAATTTACAGATTTTTTTTAAATGAGAAA 3140

Qy   3159 CACACAGA---AGAAGCTACCTCAGGTG-TTTTTTACCTCAGCCACCTTGCTCTTGTGTTTC 3214
                  |||||| |||| ||||||| ||||||||||||||||||||||||||||||||
Db   3141 AACACACAAATAAAAGCTACCTCCAGGTGTTTTTTACCTCAGCCACCTTGCTCTTGTGTTTC 3200

Qy   3215 CCTTAGAGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATGA 3274
                  ||||||| ||||||||||||||||| || ||||| |||||||||||||||||||||||||||
Db   3201 CCTTAGAGATTTTGTAAAACTGATAGTTGGAGCA-TTTTTTATTTTTTTAATAAAAATGA 3259

Qy   3275 GTT 3277
                  |||
Db   3260 GTT 3262

RESULT 3
AK044984
LOCUS   AK044984   2964 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130017M24 product:HEPATOCELLULAR CARCINOMA
            AUTOANTIGEN homolog [Homo sapiens], full insert sequence.
ACCESSION   AK044984
VERSION   AK044984.1 GI:26090647
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
PUBMED   10349253
MEDLINE   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

```
           /clone_lib="RIKEN full-length enriched mouse cDNA library"
           /dev_stage="9.5 days embryo"
misc_feature
           1..2964
           /note="HEPATOCELLULAR CARCINOMA AUTOANTIGEN homolog [Homo
           sapiens] (SPTR|Q9Y6M; evidence: FASTY, 88.3%ID,
           100%length, match=1775)"
ORIGIN

Query Match     54.6%;  Score 1862.8;  DB 11;  Length 2964;
Best Local Similarity   81.9%;  Pred. No. 0;
Matches 2441;  Conservative  0;  Mismatches 452;  Indels  87;  Gaps  22;

Qy      247 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTA 306
            ||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||
Db       30 GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGACTACTCAGTCTCTAAAAGCTA 89

Qy      307 AGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCACCTGCAGTGGGAGTGTTGGAT 366
            ||||||||||||| ||||||||||||||||||||| ||||||||||||||||||
Db       90 AGGAGCAGGAGAATCCAGATTCGGAACATCCGAGTCCACCTGCAGTGGGAGTGTTCGAT 149

Qy      367 GGACTTTTGGCTCAATATGGGACAGTGGAAGATGTGGAACAAGTCAACACAGACACAGAA 426
            |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db      150 GGGCGTGTTGGCTGAATATGGGACAGTGGAGAACGTGGAGCAAGTCAACACAGATACAGA 209

Qy      427 ACCGCCGTTGTCAACGTCACATATGCACATGGACTTCTCTACAGACTACTCAGTCTCTCG 486
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      210 ACTGCCGTTGTCAACGTCACCTATCAAGCAAAGTCTAGCTATTGAGAAG 269

Qy      487 CTAAGCGGGCATCAGTTTGAGAACTACTTCCTTCAAGATTTCTACATCCCGATGAAGAG 546
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      270 CTCAGTGGGCATCAGTTTGAGGACTACTTCCTTCAAGATTTCCACATCCCGATGAAGAG 329

Qy      547 GTGGAGCTCCCTTCGCCCCCTCAGCGAGCCCCAGCGTGGGACCACTCCTTCCCGGGACCAA 606
            ||||||||||||||||| |||||||||||||||||| 
Db      330 GTGGAGCTCCCTTCACCCCCTCATCG----------------TGCCCCGGGACAA 368

Qy      607 GGCCACAGCCCCTGGGGGCACTTCTTCAGGCGAGCAGACAGATTGATTCCCGTGCGGATCCTG 666
            |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      369 GGCCACGGCCCCTGGGGGCACTTCTTTGTGCCTGCAGAGATTGATTCCCGTGCGGATCCTG 428

Qy      667 GTCCCCAGTTTGTTCGTTGCGTGGCATCATCGGAAAGAGGCTTGACCATAAAGAACATC 726
            |||||||||||||||||||||||||||||||| |||||||||||||
Db      429 GTCCCCAGTTTGTTCGTTGCGTGGCATCATCGGAAAGAGGCTTGACCATAAAGAACATC 488

Qy      727 ACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAG 786
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      489 ACTAAGCAGACCCAGTCCCGGGTAGACATCCACAAGAGGAGAACTCTGGGGCTGCAGAG 548

Qy      787 AAGCCTGTCACCATCCATGCCAAGATTCTGAAGCATGCCGATGATTCTT 846
            |||||||||||||||| ||||||||||||||||||||||||||
Db      549 AAGCCTGTCACAATGGCTTCGTTGGAAGACATCTGAAGCATGCCGATGATTCTT 608

Qy      847 GAAATCATCGAAAGAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCTCTGAAATC 906
            |||||||||||||||||||||||||||||||||||
Db      609 GAGATTATGCAAAAGAGAGCTGATGAGACAGAGTTCCTCTGAAATC 668

Qy      907 TTGGCACACAATGGCTTCGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAA 966
            ||||||||||||||| |||||||||||||||||||||||||||||||
Db      669 CTGGCCCCACAATGGCTTCGTTGGAAGACTCATTGCCAAAGAAGGCAGAAACCTGAAGAA 728

Qy      967 ATTGAACATGAAACAAGGGACCAAGAAGCATAACAATCTCATCTTTGAAGATTTGAGCATATAC 1026
            |||||||||| ||||||||||||||||||||||||||||||||||||||
Db      729 ATAGAACATGAGACAGAACCAAGAAGCATAACAATCTCATCTTTGCAGATTTGAGCATTTAT 788

Qy     1027 AACCCGGAAAGAACCATCACTGTGAAGGCGCACAGTTGAGCCTGTCCGTGCCTGAGATA 1086
            |||||||||||||||||||||||||||||||||||||||||||||||
Db      789 AACCCCGAGAGAACCATCACGTGAAGGCGCACAGTTGAGCCTGTCCGTGCCTGAGATA 848

Qy     1087 GAGATTATGGAAGAAGCTGCGTGAGCGTTTGAAAATGATAGTGCTGGCTGGCCGTGTTAACCAACAA 1146
            |||||||||||||||||||||||||||||||||||||||||||||||
Db      849 GAGATTATGGAAGAAGCTCCGAGAGCGTTTGAGAGACGACATGCTGGCCGTTAACCAACAA 908

Qy     1147 GCCAATCTGATCCCAGGGGTTGAACCTCAGCGCACTTGGCATCTTTCACAGGACTGTCC 1206
            |||||||||||||||
```

```
Db      909 GCCAATCTGATCCGCAGGGCTAAACCTCAGTGCACTGGCACTCCCGCGCCCCGCGAGCTCCCCCGTACCACCCTTC 968
            |||||||||||| || || |||||||||||||||| |||||||||| 

Qy     1207 GTGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCCGCACCACCACCAGTTGCCGCATCTTTCGGACTGGACTGTCT 1266
            ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      969 GTGCTTCTCCACCAGCAGGGCCCCGTGGAGTTCCCCGCAGTCCCCCATCCTCCCATCCACCCGCTTT 1028

Qy     1267 ACTACCCACTCCGGATACTTCTCCAGCGCTGTACCCCATCACCAGTTGGCCGTTCCCGTTCCCG 1326
            ||||||||||||||||||||||| || ||||||||||||||||||||||||
Db     1029 GCTACCCACTCCGGATACTTCTCCAGTCGTGTACCCTCATCACCATTCGGCCCATTCCCA 1088

Qy     1327 CATCATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCAACCCAGGCTGTG 1386
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1089 CATCATCACTCCTACCCAGACAGGAGATGTAAGTCTCTTCGTCCCAACCCAGGCTGTG 1148

Qy     1387 GGCGCCATCATCGGGAGAAGGGGCACACATCAAACAGTCTGGCGAGATTGCGGAGCC 1446
            ||| ||||||||| |||||||||| ||||||||||||||||||||| 
Db     1149 GGTGCTATCATCGGGAGAAGGGGCACACATCAAACAGTCTGCTCGATTTGCTGTGTGCC 1208

Qy     1447 TCTATCAAGATGCCCCTGCCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCACC 1506
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     1209 TCCATCAAGATTGCTCCAGCGAAGGTCCAGATGTCAGTGTGGTCATCATCACT 1268

Qy     1507 GGGCCACCGGAGCCCAGTTCAAGGCCCAGGAGACGGATCTTTGGGAAACTGAAAGAGGAA 1566
            ||||||||||||||||||||||||||||||||||||||||
Db     1269 GGTCCTCCTGAAGCCCAGTTTAAGGCTCAGGAGACGGATCTTTGGGAAACTGAAGGAAGAA 1328

Qy     1567 AACTTCTTAACCCAAAGAAGAAGTGAAGCTGGCAAAGGTTGGCAAGACCGTGAACGAACTGCAGAACTTAACC 1626
            ||||||||||||||||||||||||||||||||||||||
Db     1329 AACTTCTTTAATCCAAAGAAGAAGTGAAGCTGGCAAAGGTGGCAAAACCGTGAACGAACTTGACA 1388

Qy     1627 ACAGCTGGCCGGGTGATTGGCAAAGGTTGGCAAGACCGTGAACGAACTGCAGAACTTGACA 1686
            |||||||||||||||||||||||||||||||||||||||||
Db     1389 ACCGCTGGCCGGGTGATTGGCAAAGGTGGCAAAACCGTGAACGAACTGACA 1448

Qy     1687 AGTGCAGAGTCATCGTGCCTCTGTGACCAGCCCAGATGAAAATGAGGAAGTGATCGTC 1746
            |||||||||||||||||||||||||||||||||||||||||
Db     1449 AGTGCAGAGTTATCGTGCCTCTGTGACCAGCCCAGATGAAATGAGGAAGTGATCGTC 1508

Qy     1747 AGAATTATCGGGACACTTCTTTGCTAGCCAGATCAGCGCACGCAAGATCAGGGAAATTGTA 1806
            ||||||||||||||||||||||| ||||||||||||||||||
Db     1509 AGAATTATCGGGACATTTTTTCGTGCTAGCCAGATCAGCGCACGCAAGATCAGGGAAATTGTA 1568

Qy     1807 CAACAGTGAAGCAGCAGCAGGAGCAGAGAAATACCCTCAGGAGTGCGCCTCAGCGGCAGCAAG 1866
            |||||||||| ||||||||||||||| ||||||||||||||||||||
Db     1569 CAGCAGTGAAGCGAGCAGCAGGAGCAGAGATACCCTCAGGAGTGCGCCCACAGCGGCAGCAAG 1628

Qy     1867 TGAGGCTCCACAGGACGCACAGCACGGATGAATGTAGCCTGCAACACCTGACA 1926
            |||||||| ||||| || |||||||||| |||||||||
Db     1629 TGAGGCTCCACAG--CACCAGCAAGCAACCGATGAATGTAGCCTCCCAACACCTGACA 1686

Qy     1927 GAATGAGACCAAAGCAGCCAGCCAGATGCGGAGCAAACCAAAGACCAT-CTGAGGAATG 1985
            |||||||||||||||||| ||||||||||||||||||| ||||
Db     1687 G-ATGAGACCAAAC---AGCCAGCAGATAGGGGAGCATCCCAAAGACTGTGTTCAAGGAGTG 1742

Qy     1986 AGAAGTGCTGCGGAGCGGCCAGGAGCTGCCGAGGGACTCTGCCGAGGCCCTGGAACCCCAGGGGCCGAGG 2045
            |||||||||||| || |||||||||||||||||| 
Db     1743 CGCAGTGTGCAGAGCAGGCAGGAGCTGCGCTGC-----AGACCTCTACACATCTGGGATCCTGGCAGG 1797

Qy     2046 AGGGGCGGGGAAGGTCAGCGCCCAGGTTGCCAGAACCACCG--AGCCCCGGCTCCCGCCCCC 2103
            |||||||||||| |||||||||||| || ||||||||| 
Db     1798 AGGGCACAGGGAAG---GCCAGGGTGTCCCAGAAACACCGCTTGCCCTGGCTCCCCAGCTTCC 1854

Qy     2104 CAGGGGCTTCTGCAGGCTT--CAGCCATCCACTTCACCATCCGGATCCTCCCTGAAC 2161
            |||||||||||||||||| || ||| |||||||||||||| 
Db     1855 CCTGGCTTCTGCAGGCATAACAGCACCATCGCCCATCGCCCAACTCAGAT-TCTCCCTCAGT 1913

Qy     2162 TCCCACGACGTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCA 2221
            ||||||||||||||||||||||||||||||||||||||
Db     1914 TCCCACGACGTATCCCTTTCGGTTGACTAACATAGGTGAACTAACATGCTCAAAGCAAAAGCA 1973

Qy     2222 AAATGCACACCC----TTTTCTGTGGCCAAATCGTCTCTGTACATGTCTGTATATTAG 2277
            |||||||| ||
Db     2222
```

```
VERSION     BQ691878.1  GI:21817194
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 874)
  AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     NIH-MGC http://mgc.nci.nih.gov/.
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2362  row: h  column: 03
            High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1..874
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6207026"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match          24.7%;  Score 842.2;  DB 13;  Length 874;
Best Local Similarity 99.4%;  Pred. No. 2.3e-154;
Matches  856;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;
```

```
Qy   1824 GGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGGCGCAGCAAGTGAGGCTCCACAGGCA 1883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGGCGCAGCAAGTGAGGCTCCACAGGCA   60

Qy   1884 CCAGCAAAACAACGGATGAATGTAGCCCTTCCACACCTGACAGATGAGACCAAACGCA 1943
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     61 CCAGCAAAACAACGGATGAATGTAGCCCTTCCACACCTGACAGATGAATGAGACCAAACGCA  120

Qy   1944 GCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGCGAGGGCG 2003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGCGAGGGCG  180

Qy   2004 CCAGGGACTCTGCGAGGGCCTGAGGGGCCGAGGGGGCCGAGGGGGCCGGGAAGGTCAG 2063
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    181 CCAGGGACTCTGCGAGGGCCTGAGAACCCCGAGGGGGCCGAGGGGGCCGGGAAGGTCAG  240

Qy   2064 CCAGGTTTGCCAGAACCACCGAGCCCCGCCTCGCCCGCCTCTCGAACTCCCAGGGCTTCA 2123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCAGGTTTGCCAGAACCACCGAGCCCCGCCTCGCCCGCCTCTCGAACTCCCAGGGCTTCA  300

Qy   2124 GCCATCCACTTCACCATCCACTCCGGATCTCTCCTGAACTCCCACGACGCTATCCCTTTA 2183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCATCCACTTCACCATCCACTCCGGATCTCTCCTGAACTCCCACGACGCTATCCCTTTA  360

Qy   2184 GTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGCACACCCTTTTTCTGTG 2243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGCACACCCTTTTTCTGTG  420

Qy   2244 GCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGATGTTAAGATATGTGGC 2303
```

```
Db   1974 AAATTCCTAGCCGTTTCTTTGTTGTGGAAAGTTGTCTCTGTACATGTATGTACATATCAG 2033
Qy   2278 AAAGGGAA-GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGGTGCCTGCAGCGGTAAT 2336
Db   2034 AAGGGGAAGAGATGTTAAGAAATGTGGCCTGTGGGGTTACACAGGGGTGCCTGCAGCGTTAAT 2093
Qy   2337 ATATTTAGAAATAATATATCAAATAACTCCAATTTTAATCAATTATTAA 2396
Db   2094 ATATTTAGAAATAATATATCAAATAACTCCACTAACAACAATTTTACTCAGTGATTAA 2153
Qy   2397 TTTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGATAAAGTCTTTG 2456
Db   2154 TTTTGTTTTCTTTTGAAGAGAAAGCAGGCTTTTCTAGACTTCAAGAGAATAAAGTATTGG 2213
Qy   2457 G-GAGGTCTCACGGTGTAGAGAGAGCTTGAGGC-CACCCGCACAAAATTCACCCAGAG 2514
Db   2214 GTGAGGTCCACTGGTATAGAAAGGAACTTGAGCCACCCACACAAAAATTCACTTGGAG 2273
Qy   2515 GGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAA 2574
Db   2274 GGAAATCTTGTC-AAGGACACTTAATGGCAGTTGTGGGTCACCTGTGTCTGTCAACAGAA 2332
Qy   2575 GGGATACCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCAGTGCCTGTCTAGTCATACA 2634
Db   2333 GGGATACCGTCCCTGAGGAGGGGGCCCTAGGCCCTGGTCACCTGTTTCTCCCACACACCCA 2392
Qy   2635 CCCATTTCTCTT----------------TGCTTCACAGGTTTAAACTGGTTTTTT 2674
Db   2393 TTGCTCAGTTGCACAGGTTTTAAACTGATGATTTTGGTTTGTTTTGTTTTGTTTTTT 2452
Qy   2675 GCATACTGCTATATAATTCTCTGTCTCTCTGTTATCTCTCCCCTCCCCTCCCC 2734
Db   2453 GCATACTGCTATATAGTTCGTTCTCTC-----CTTATCTCCCTCCTAACCTTCCTCT 2506
Qy   2735 TTCTTCTCCATCCCATTCTTTTGAATTTCCTCATCCCTCCAATCCGTATCTAC 2794
Db   2507 GTCTTCTTTACCCAAAATTCTTTGAGTCCCTTCGTTCTTTTTCTGCTCCCCGTATCTAT 2566
Qy   2795 GCACC---CCCCCCCCCCCCAGGCAAAGCAGTGCTCTGAGTAGTCACATCACACAAAGGAA 2851
Db   2567 GCACCCCTCCCCTCCCCCCAGGCAAAGCAGTGCTCTGAATATCATCACACACAAGGAA 2626
Qy   2852 CAAAAGCGGAACACACAGACCACCAGCCAGCCTCAACTTACACTTGGTACTCAAAAGAACAAGAGT 2911
Db   2627 CAAATGCGAACACACAGACCACCAGCCAGCCTCAGCTTACACTCGGTTACTCAAGCGAACAAGAGT 2686
Qy   2912 CAATGGTACTTGTCCTAGCGTTTGGAAGAGGAAAAACAGGAGACCCACCAAATAAACCAAT 2971
Db   2687 CGGCGGTACTTGTTCCTAGC-ATTCGGAAGAGGAAAGCAGGAGATCCCAGCAAATAAACCAGC 2745
Qy   2972 CAACCAAACAAAGAAAAAATTCCACATGAAAGAATGAATGTATTTTGTCTTTTGCATTTTGG 3031
Db   2746 CAGCCAAACAAAG-AAAAATCCCACAAGAAAGAAAATGGA----CTCTCTCCATCTTGG 2800
Qy   3032 TGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTTCTTTTAAAAAATGTCGAGG 3091
Db   2801 TATACAAGCCATCAACATTAGTGAAATTCTTTTCTTTTCTTTTAAAATTTAAAGTGCAGGA 2860
Qy   3092 AAAGTAGAAATTTACCAAGGTTGTTGGCCCAGGCGTTAAATTCACAGATTTTTTTTAACG 3151
Db   2861 AAATAGCAATTTATACGAGGTTGTTGGCCCAGGGCGTTAAATTTACAGATTTTTTTTTAAA 2920
Qy   3152 AGAAAAACACACAGA--AGAAGCTACCTCAGGTGTTTT 3188
Db   2921 TGAGAGAAAACACACAAATAAAAAGCTACCTACCTCAGGTGTTTT 2960
```

```
RESULT 4
BQ691878
LOCUS       BQ691878     874 bp    mRNA    linear    EST 15-JUL-2002
DEFINITION  AGENCOURT_8034698 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6207026
            5', mRNA sequence.
ACCESSION   BQ691878
```

```
Query Match                 24.6%;  Score 841;  DB 12;  Length 905;
Best Local Similarity  97.1%;  Pred. No. 3.9e-154;
Matches  878;  Conservative  0;  Mismatches  21;  Indels  5;  Gaps  2;

Qy   492 CGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGATGAAGAAGAGGTGAG 551
         ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     1 CGGGCATCAGTTTGAGAACTACTCCTTCAGATTCCTACATCCCGATGAAGAAGAGGTGAG 60

Qy   552 CTCCCCTTCGCCCCCTCAGCCAGCCCAGCGTGGGACCACTCTTCCGGGAGCAAGGCCA 611
         || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTCCCCTTCGCCCCCTCAGCCAGCCCAGCGTGGGACCACTCTTCCGGGAGCAAGGCCA 120

Qy   612 CGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGATCCTGGTCCC 671
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCGCTGCGATCCTGGTCCC 180

Qy   672 CACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGGCTTGACCATAAAGAACATCACTAA 731
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGGCTTGACCATAAAGAACATCACTAA 240

Qy   732 GCAGACCCAGTCCGGGTAGATATCCGGAGAGAAAAGAACTCGAGCTGCGAGAGAAGCC 791
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCAGACCCAGTCCGGGTAGATATCCGGAGAGAAAAGAACTCGAGCTGCGAGAGAAGCC 300

Qy   792 TGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGAAGCATGCCGATGATTCTTGAAAT 851
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGTCACCATCCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGAAGCATGCCGATGATTCTTGAAAT 360

Qy   852 CATGCAGAAGAGGCAGATGAGGACCAAACTAGCCGAAGAGAATTCCTCGAAAATCTTGGC 911
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CATGCAGAAGAGGCAGATGAGGACCAAACTAGCCGAAGAGAATTCCTCGAAAATCTTGGC 420

Qy   912 ACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATTTGAAGAAATTGA 971
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGAAGGCAGAAATTTGAAGAAATTGA 480

Qy   972 ACATGAAACAGGACCAAGATAACAATCAATCATCTTGAGCGTTGAGCATATACACACCC 1031
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACATGAAACAGGACCAAGATAACAATCAATCATCTTGAGCGTTGAGCATATACACACCC 540

Qy  1032 GGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGCGCAGTGCTGAGATAGAGAT 1091
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGCGCAGTGCTGAGATAGAGAT 600

Qy  1092 TATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAA 1151
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAA 660

Qy  1152 TCTGATCCCAGGGTTGAACCTCAGCGCACTTGCCATCTTTCAACAGGACTGTCCGTGCT 1211
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCTGATCCCAGGGTTGAACCTCAGCGCACTTGCCATCTTTCAACAGGACTGTCCGTGCT 720

Qy  1212 ATCTCCACCAGCAGGGCCCCGGAGTCCCCCGTGCCCCCGTGCCCCCTACCACCCCTTCACTAC 1271
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ATCTCCACCAGCAGGGCCCCGGAGTCCCCCGTGCCCCCGTGCCCCCTACCACCCCTTCACTAC 780

Qy  1272 CCACTCCGGATACTTCTCCGACCTGTACCCCCATCACCAGTTT--GGCCCGTTCCCGGCAT 1329
         |||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db   781 CCACTCCGGATACTTCTCCGACCTGTACCCCCATCACCAGTTTTGGCCCGTTCCCGGCAT 840

Qy  1330 CATCACTCTTATCCAGAGGAGGAGATTGTG---AATCTTCATCCCAACCCAGGCTGTG 1386
         |||||||||||||||||||||||||||||   ||||||||||||||||| ||||||||
Db   841 CATCACTCTTATCCACGAGCAGGAGAGATTGTGAAATCTTCATCCCAANCCCAGGCCTG 900

Qy  1387 GGGG 1390
         ||||
Db   901 TGGG 904
```

RESULT 6
BQ221568
LOCUS   BQ221568   851 bp   mRNA   linear   EST 02-MAY-2002

```
Db   421 CAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAGATT 480

Qy   527 CCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCCGTGGGG 586
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CCTACATCCGGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCCGTGGGG 540

Qy   587 ACCACTCTTCCGGGAGCAAGCCACGCGCCCCTGGGGCACTTCTCAGGCAGACAGATTG 646
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACCACTCTTCCGGGAGCAAGCCACGCGCCCCTGGGGCACTTCTCAGGCAGACAGATTG 600

Qy   647 ATTTCCGCTGCGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGG 706
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATTTCCGCTGCGGATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGG 660

Qy   707 GCTTGACCATAAAGAACATCACTAAGCAGACACCCAGTCCCGGTAGATATCCATAGAAAG 766
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GCTTGACCATAAAGAACATCACTAAGCAGACACCCAGTCCCGGTAGATATCCATAGAAAAG 720

Qy   767 AGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTG 826
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AGAACTCTGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGGACTTCTG 780

Qy   827 AAGCATGCCGCATGATTCTTGAAATCATGCAGAAAGAGCAGAGATGAGACCAAAC-TAGCC 885
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   781 AAGCATGCCGCATGATTCCTGAATCCTGCGAATCCTGCGAAGAGGCAGAGATGAGACCAACTTAGCC 840

Qy   886 GAAGAGATTC 895
             ||||||||||
Db   841 GAAGAGATTC 850
```

RESULT 7
BM561242
LOCUS   BM561242   999 bp   mRNA   linear   EST 20-FEB-2002
DEFINITION AGENCOURT_6557353 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739550
            5', mRNA sequence.
ACCESSION   BM561242
VERSION   BM561242.1 GI:11806337
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 999)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMl2753 row: a   column: 23
            High quality sequence stop: 679.
FEATURES   Location/Qualifiers
     source   1..999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5739550"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

DEFINITION AGENCOURT_7559207 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6046477
            5', mRNA sequence.
ACCESSION   BQ221568
VERSION   BQ221568.1 GI:20402968
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 851)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMl3292 row: f   column: 14
            High quality sequence stop: 637.
FEATURES   Location/Qualifiers
     source   1..851
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6046477"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."
ORIGIN

Query Match   24.3%; Score 830; DB 13; Length 851;
Best Local Similarity   99.3%; Pred. No. 5.6e-152;
Matches 844; Conservative   0; Mismatches   0; Indels   5; Gaps   1;

Qy   47 GCGGGCTCTCGGGGGAAGAGACGGATGATGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCG 106
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCGGGCTCTCGGGGGAAGAGACGGATGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCG 60

Qy   107 CCGTCACCGCGACGACCTCCGGCAGCTCTTTGGGACAGCAGGAAGCTGCCCCTGGCGGGAC 166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGTCACCGCGACGACCTCCGGCAGCTCTTTGGGACAGCAGGAAGCTGCCCCTGGCGGGAC 120

Qy   167 AGGTCCTGCTGAAGTCCGGCTACGCCTTCGTCGGACTACCCCGACCAGAACTGGGCCATCC 226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGTCCTGCTGAAGTCCGGCTACGCCTTCGTCGGACTACCCCGACCAGAACTGGGCCATCC 180

Qy   227 GCGCCATCGAGACCCCTCTCGGGTAAAGTGGGAATTGCATGGGAAAATCATGGAAGTTGATT 286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCGCCATCGAGACCCCTCTCGGGTAAAGTGGGAATTGCATGGGAAAATCATGGAAGTTGATT 240

Qy   287 ACTCAGTCTCTAAAAAGCTAAGGAGCAGGAGAAATTCGAACATCCCTCCTCACC 346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACTCAGTCTCTAAAAAGCTAAGGAGCAGGAGAAATTCGAACATCCCTCCTCACC 300

Qy   347 TGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAAC 406
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGCAGTGGGAGGTGTTGGATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAAC 360

Qy   407 AAGTCAACACAGACACAGAAACCGCCGGTTGTCAACGTCACGTCACATATGCAACGACAAGAAGAAG 466
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGTCAACACAGACACAGAAACCGCCGGTTGTCAACGTCACATATGCAACGACAAGAGAAGAAG 420

Qy   467 CAAAAATAGCCATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAGATT 526
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Lauber,J., Bahr,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Lauber,J., Bahr,A., Mewes,H.W., Weil,B., Amid,C., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp781G0234) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp781G0234"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="781 (synonym: hlcc4)"
                     /note="Vector: pSport1_sfi; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN
```

```
Query Match       23.7%;   Score 809; DB 13; Length 832;
Best Local Similarity  99.9%;  Pred. No. 7.3e-148;
Matches 820;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;
```

```
Qy   1732 GAGGAAGTGATCGTCAGATTATCGGCACTTCTTGCTCTTTGTTAGCCAGACTGCACAGCGCAG 1791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13 GAGGAAGTGATCGTCAGATTATCGGCACTTCTTGCTCTTTGTTAGCCAGACTGCACAGCGCAG 72

Qy   1792 ATCAGGGAAATTGTACAACAGGTGAAGCAGTGAAGCAGAGAGCAGGAGAGAAATACCCTCAGGGAGTCGCC 1851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 ATCAGGGAAATTGTACAACAGGTGAAGCAGTGAAGCAGAGAGCAGGAGAGAAATACCCTCAGGGAGTCGCC 132

Qy   1852 TCACAGCGCAGCAAGTGAGGCTCCCACAGCCACCAGGCACCAGGCAAAACAACGGATGAATGAATGTAGCCC 1911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 TCACAGCGCAGCAAGTGAGGCTCCCACAGCCACCAGGCACCAGGCAAAACAACGGATGAATGAATGTAGCCC 192

Qy   1912 TTCCAACACCTGACAGAATGAGACAAGCGCAAGATCAGGGAAATTG 1971
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    193 TTCCAACACCTGACAGAATGAGACAAGCGCAAGATCAGGGAAATTG 252

Qy   1972 CCATCTGAGGAATGAGAAGTCTGCGGACAGGAGTCTGCGGAGGACTCTGCGGAGGAAC 2031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CCATCTGAGGAATGAGAAGTCTGCGGACAGGAGTCTGCGGAGGACTCTGCGGAGGAAC 312

Qy   2032 CCCAGGGGCCGAGGAGGGGGGCGGGGGCGGGGGCCCCAGGAGTTCAGCAGGCTTCAGCGGCCCCCG 2091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 CCCAGGGGCCGAGGAGGGGGGCGGGGGCGGGGGCCCCAGGAGTTCAGCAGGCTTCAGCGGCCCCCG 372

Qy   2092 CCTCCCGCCCCGCCCCGGGCTTCTGCAGGCTTCAGCGGATC 2151
          |||||||||||||||||||||||||||||||||||||||||
Db    373 CCTCCCGCCCCGCCCCGGGCTTCTGCAGGCTTCAGCGGATC 432

Qy   2152 TCTTCCTGAACTCCCAGCAGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCA 2211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 TCTTCCTGAACTCCCAGCACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCA 492

Qy   2212 AAGCCAAGCAAAATGCACACCCTTTTCTGTGCGCAAATGCTCTGTGACATGTGTGTACA 2271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   493 AAGCCAAGCAAATGCACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTGTACA 552
Qy   2272 TATTAGAAAGGGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGGG 2331
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 TATTAGAAAGGGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGGG 612
Qy   2332 GTAATATATTTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATT 2391
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   613 GTAATATATTTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATT 672
Qy   2392 ATTAATTTTTTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGT 2451
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   673 ATTAATTTTTTTTTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGT 732
Qy   2452 CTTTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCA 2511
            ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   733 CTTTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCAC-AAATTCACCCA 791
Qy   2512 GAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGA 2552
            ||||||||||||||||||||||||||||||||||||||||
Db   792 GAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGA 832

RESULT 9
BM806132
LOCUS   BM806132   1089 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION AGENCOURT_6553922 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555652
            5', mRNA sequence.
ACCESSION   BM806132
VERSION   BM806132.1 GI:19122955
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12276 row: c column: 13
            High quality sequence stop: 688.
FEATURES   Location/Qualifiers
     source   1..1089
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5555652"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "
ORIGIN

Query Match   23.5%; Score 803.4; DB 12; Length 1089;
Best Local Similarity   94.0%; Pred. No. 8.8e-147;
Matches 936; Conservative   0; Mismatches 42; Indels   18; Gaps   9;

Qy   201 CTACCCGCACCGACCAACTGGGCCGCCATCCGCGCCATCCGCGACCCCTCTCCGGGTAAAGTGGAATT 260
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   1 CTACCCGCACCGACCAACTGGGCCGCCATCCGCGCCATCCGCGACCCCTCTCCGGGTAAAGTGGAATT 60
Qy   261 GCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGGACGCAGGAAAAT 320

Db   61   GCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGGACGCAGGAAAAT 120
Qy   321 TCAGATTGAAAACATCCCTGCTGCGGTGGGAGGTGTTGGATGGACTTTGGGTGTTGGATGGACTTTGGGTCA 380
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCAGATTCGAAACATCCCTGCTGCGGTGGGAGGTGTTGGATGGACTTTGGGTCA 180
Qy   381 ATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACACAGACACGCCGTTGTCAA 440
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACACAGACACGCCGTTGTCAA 240
Qy   441 CGTCACATATGCAACAGAGAGAAGCAAAAATAGCCATGAGAGAAGAAGCTAAGCGGGCATCA 500
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CGTCACATATGCAACAGAGAGAAGCAAAAATAGCCATGAGAGAAGAAGCTAAGCGGGCATCA 300
Qy   501 GTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTC 560
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGCTCCCCTTC 360
Qy   561 GCCCCCTCAGCGGACGCCAGCGTGGGGACCCACTCTTCCCGGGTGCGGATCCTGCGTGTCCCCACCCAGTT 620
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GCCCCCTCAGCGGACGCCAGCGTGGGGACCCACTCTTCCCGGGTGCGGATCCTGCGTGTCCCCACCCAGTT 420
Qy   621 GGGCACTTCTCAGCCAGACAGATTGATTTCCCGCTGCGGAAATCTGAAATCATGCAGAA 680
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGGCACTTCTCAGCCAGACAGATTGATTTCCCGCTGCGGAAATCTGAAATCATGCAGAA 480
Qy   681 TGTTGGTGCCATCCATCGGAAAGGAGGGCTTGACCATAAACATCACTAAGCAGACCCA 740
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGTTGGTGCCATCCATCGGAAAGGAGGGCTTGACCATAAACATCACTAAGCAGACCCA 540
Qy   741 GTCCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTCGCAGAGAAGCCTGTCACCAT 800
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GTCCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTCGCAGAGAAGCCTGTCACCAT 600
Qy   801 CCATGCCACCCCGAGGGGACTTCTGAAGCATGCCCATGATTCTTGAAATCATGCAGAA 860
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CCATGCCACCCCGAGGGGACTTCTGAAGCATGCCCATGATTCTTGAAATCATGCAGAA 660
Qy   861 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCACACAATGG 920
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAANATCTTGGCACACAATGG 720
Qy   921 --CTTGGTTGGAGACTGATTGGAAAAGAAGGCAGAAATTGAACATGAA 978
              |||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GCTTGGTTGGAGACTGATTGGAAAAGAAGGCAGAAATTGAACATGTA 780
Qy   979 ACA-GGGACCAAGATAACAATCTCATCTTT-GCAGGATTTGAGCATATACAACCCGG-AA 1035
            ||| |||||||||||||||||||||||||| |||||||||||||||||||||||| ||
Db   781 ACAGGGGACCAAGATAACCATCTCATCTTTGCAGGATTTGAGCATATACCACCCGGAA 840
Qy   1036 AGAACCATCACTG-TGAAGGGCACAG-TTGAGGCCGTGCCCAGTGCCAGTGATAGAGATTA 1093
            ||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||
Db   841 AGAACCATCACTGTTGAAGGGCACAGTTTGAGGCCGTGCCCAGTTGCCCATGCCGAAGATAGAGAT 900
Qy   1094 TGAAGAAGCTGC-----GTGAGGGCCTTTGAAAATGATATGCTGG--CTGTTAACCAAC 1144
            |||| | |||| |||||||||||||||||||||||| ||||||||||
Db   901 TATGGAAGAAACCTGGCCGAGGGGCCCTTTGAAAATGAAATGCTGGCGTCGTTTAACCAAC 960
Qy   1145 AAGCC--AATCTGATCCCCAGGGTTGAACCTCAGCGC 1178
            ||||| |||||||||||||||||||||||||||
Db   961 AAACCCAATCTTGATCCCCAGGGTTGAACCTCCACGC 996

RESULT 10
BQ212500
LOCUS   BQ212500   836 bp   mRNA   linear   EST 02-MAY-2002
DEFINITION AGENCOURT_7675661 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6095809
            5', mRNA sequence.
ACCESSION   BQ212500
VERSION   BQ212500.1 GI:20392782
KEYWORDS   EST.
SOURCE   Homo sapiens (human)

ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 836)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC/DCTD/DTP
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13368 row: n column: 02
          High quality sequence stop: 670.
FEATURES   Location/Qualifiers
     source   1..836
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:6095809"
                   /tissue_type="melanotic melanoma"
                   /lab_host="DH10B (phage-resistant)"
                   /clone_lib="NIH_MGC_72"
                   /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                   Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                   Average insert size 2 kb. Library constructed by Life
                   Technologies."

ORIGIN

Query Match   23.4%; Score 799.2; DB 13; Length 836;
Best Local Similarity   98.7%; Pred. No. 6e-146;
Matches 826; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

Qy 2168 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCAAGCAAGCAAATGC 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAGCAAGCAAGCAAATGC 60

Qy 2228 ACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAGGGGAGA 2287
          |||||||||||||||||||||||||||||||||||||||| |||||||
Db   61 ACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGGGAAGA 120

Qy 2288 TGTTAAGATATGTGGCCTGTGGGTTACACAGAGGGTGCCTGCAGCCGGTAATATTTTAGAA 2347
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   121 TGTTAAGATATGTGGCCTGTGGGTTACACAGAGGGTGCCTGCAGCCGGTAATATATTTAGAA 180

Qy 2348 ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTCT 2407
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   181 ATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTAATCAATTTTTTTTTTTCT 240

Qy 2408 TTTTAAAGAGAAAGCAGGAGCTTTGAGGCTTTCTAGACTTAAGAATAAAGTCTTGGGAGGTCTCAC 2467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTTTAAAGAGAAAGCAGGAGCTTTGAGGCTTTCTAGACTTAAAGAATAAAGTCTTGGGAGGTCTCAC 300

Qy 2468 GGTGTAGAGAGGAGCTTTGAGGCCACACCCGCACAAAATTCACCCAGGAGGGAAAATCTCGTCG 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGTGTAGAGAGGAGCTTTGAGGCCACACCCGCACAAAATTCACCCAGGAGGGAAAATCTCGTCG 360

Qy 2528 GAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATACCGTCTC 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATACCGTCTC 420

Qy 2588 CTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTCAGCTCATCACCCATTTCTCTT 2647
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   421 CTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTCAGCTCATCACACCCATTTCTCTTT 480

Qy 2648 GCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATAAATTCTCTGTCTCTCTCTG 2707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAAATTCTCTGTCTCTCTCTG 540

Qy 2708 TTTATCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCTTCTTCTCCATCTCCATTCTTTGAATTTCCTC 2767
          |||||||||||||||||||||||||||||||||||||||||||
Db   541 TTTATCTCTCCCTCCCTCCCTCCCTCCCTCCTTCTTCTCCATCTCCATTCTTTGAATTTCCTC 600

Qy 2768 ATCCCTCCATCTCAATCCGTATCACGACCCCCCCCCAGGCAAAGCAGTGCTCT 2827
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   601 ATCCCTCCATCTCAATCCGTATCACGCA-CCCCCCCCCCAGGCAAAGCAGTGCTCT 659

Qy 2828 GAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAACACCAGCCTCAACTTACAC 2887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 GAGTATCACATCACACAAAAGGAACAAAAGGCGAAACAAACAACACCAGCCTCAACTTACAC 719

Qy 2888 TTGGTTACTCAAAGAGACAAGAGTCAATGGTACTTGTCCTA-GCGTTTGGAAGAGAAA 2946
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   720 TTGGTTACTCAAAGAGACAAGAGTCAATGGTACTTGTCCTANNCGTTTGGAAGAGAAA 779

Qy 2947 ACAGGAACCCACCAAACCCAACCAATCAACCAAACAAGAAAAATTCCACCATGAAA 3003
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   780 ACAGGAACCCACCAAACCCAACCAATCAACCAAACAAGAAAAAATTCCACCATGAAA 836

RESULT 11
BU156245
LOCUS   BU156245   894 bp   mRNA   linear   EST 03-SEP-2002
DEFINITION AGENCOURT 7965128 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6162660
            5', mRNA sequence.
ACCESSION   BU156245
VERSION   BU156245.1 GI:22669777
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 894)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13516 row: o column: 13
            High quality sequence stop: 666.
FEATURES   Location/Qualifiers
     source   1..894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6162660"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."

ORIGIN

Query Match   22.3%; Score 759.8; DB 13; Length 894;
Best Local Similarity   95.5%; Pred. No. 3.1e-138;
Matches 836; Conservative 0; Mismatches 32; Indels 7; Gaps 5;

Qy 2102 CCCAGGGCTTCTGCAGGCTTCAGCCATCACCATTCACCATTCACCATCCTCGGATCTCGGATCTCCTGAAC 2161
          |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   20 CCCTCCTCACCTGCAGGCTTCAGCCATCACCATTCACCATCCACCATCCTCGGATCTCCTGAAC 79

Qy 2162 TCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCA 2221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10576 row: m column: 04
High quality sequence stop: 773.

```
FEATURES             Location/Qualifiers
     source          1..781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4707243"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_87"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match              22.1%;   Score 752.6;  DB 12;  Length 781;
Best Local Similarity    98.6%;   Pred. No. 8.1e-137;
Matches 770;  Conservative 0;  Mismatches 9;  Indels 2;  Gaps 1;
```

```
Qy 1821 GCAGGAGCAGAAATACCCTCAGGGAGTGCGCTCACGGCGCAGCAAGTGAGGCTCCCACAG 1880
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCAGGAGCAGAAATACCCTCAGGGAGTGCGCTCACGGCGCAGCAAGTGAGGCTCCCACAG 60

Qy 1881 GCACCAGCAAAACAACGGATGAATGAGCCTTCAACACTGACAGAATGAGACCAAAC 1940
        |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   61 GCACCAGCAAAACAACGGATGAATGAGCCTTCCAACACTGACAGAATGAGACCAAAC 120

Qy 1941 GCAGCCAGCCAGATCGGGAGCAGCAAACCAAAGACCATCATCTGAGGAATGAGAAGTCTGCGGAGG 2000
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCAGCCAGCCAGATCGGGAGCAGCAAACCAAAGACCATCATCTGAGGAATGAGAAGTCTGCGGAGG 180

Qy 2001 CGGCCACGGACTCTGCCGAGGCCCTGAGAACCCCAGGGCCGAGGGGGCCGGGGAAGGT 2060
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CGGCCACGGACTCTGCCGAGGCCCTGAGAACCCCAGGGCCGAGGGGGCCGGGGAAGGT 240

Qy 2061 CAGCCAGGTTTGCCAGAACCACCAGGACCCCGCCTCCCGCCCCCAGGGCTTTCTGCAGGCT 2120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CAGCCAGGTTTGCCAGAACCACCAGGACCCCGCCTCCCGCCCCCAGGGCTTCTGCAGGCT 300

Qy 2121 TCAGCCATCCACTTCACCATCCACTCCCGGATCTCTCCTGAACTCCCACGACGCTATCCCTT 2180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TCAGCCATCCACTTCACCATCCACTCCCGGATCTCTCCTGAACTCCCACGACGCTATCCCTT 360

Qy 2181 TTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCACACCCTTTTCT 2240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGCACACCCTTTTCT 420

Qy 2241 GTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGATGTTAAGATATGT 2300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGGGAAGATGTTAAGATATGT 480

Qy 2301 GGCCTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATTTAGAAATAATAATCAA 2360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GGCCTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATTTAGAAATAATAATCAA 540

Qy 2361 TAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTCTTTTAAAGAGAA 2420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTTTTTCTTTTAAAGAGAA 600

Qy 2421 GCAGCCTTTTCTAGACTTTAAAGAATACAGTCTTGGGAGGTGCTCACGGTGTAGAGAGGA 2480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GCAGCCTTTTCTAGACTTTAAAGAATACAGTCTTGGGAGGTGCTCACGGTGTAGAGAGGA 660
```

```
Db   80 TCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCA 139

Qy 2222 AAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAG 2281
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  140 AAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAG 199

Qy 2282 GGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATT 2341
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  200 GGAAGATGTTAAGATATGTGGCCTGGGTTACACAGGGTGCCTGCAGCCGGTAATATATT 259

Qy 2342 TTAGAAATAATAATCAAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTT 2401
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 TTAGAAATAATAATCAAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTTT 319

Qy 2402 TTTTCTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATACAGTCTTGGGAGG 2461
        | |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  320 TTTTTCTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATACAGTCTTGGGAGG 379

Qy 2462 TCTCACGGTGTAGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATC 2521
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  380 TCTCACGGTGTAGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAATC 439

Qy 2522 TCGTCGGAAGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGAGGGATAC 2581
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  440 TCGTCGGAAGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGAGGGATAC 499

Qy 2582 CGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGTCACTCTAGCTCTCATACACCCATT 2641
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  500 CGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCAGCTCTCATACACCCATT 559

Qy 2642 CTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTC 2701
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  560 CTCTTTGCTTCACAGGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTC 619

Qy 2702 TCTCGTTTATCTCCCCCTCCCCCCTCTTCTTCTCCCATCTCCA-TTCTTTTGAA 2760
        |||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db  620 TCTCGTTTATCTCCCCTCCCCCCTCTTCTTCTCCCATCTCCATTCTTTTCTTTTGAA 679

Qy 2761 TTTCCTCATCCCTCCATCTCAATCCGTATCCGACACCCCCCCCCCCCAGGCAAAGCA 2820
        |||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db  680 TTTCCTCATCCCTCCATCTCAATCCGTATCCGGCA--CCCCCCCCCCAGGCAAAGCA 737

Qy 2821 GTGCTCTGAGTATCACATCACAGGAACAAAAGCGAAACACACAAAACCAGCCTCAA 2880
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  738 GTGCTCTGAGTATCACATCCCCAAAAGGACAAAAGCGAAACCACCACCAGCCTCAA 797

Qy 2881 C-TTACACTTGGTTACTCAAAG-AACAAGAGTACTTGTCCTAGCGTTTTGG- 2937
        | |||||||||||||||||||| ||||||||||||||||||||||||||||
Db  798 CTTTACACTTGGTTACTCAAAAGCCAGGGTAACTTGGCCTAACCCGTTTGGG 857

Qy 2938 -AAGAGGAAAACAGGAACCCACCCAAACCAACCAAT 2971
         |||||||||||||||||||||||||||||||||||
Db  858 AAGAAGGAAAACGGAAACCCCCCAAACCAACCAACCAT 892
```

```
RESULT 12
BG575889

LOCUS       BG575889       781 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602598315F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707243 5',
            mRNA sequence.
ACCESSION   BG575889
VERSION     BG575889.1  GI:13583542
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
Qy  2481  GCTTTGAGGCCACCCGCACAAAATTCACCCAGAGAGGGAAATCTCGTCGGAAGGACACTCAC  2540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GCTTTGAGGCCACCCGCACAAAATTCACCCAGAGAGGGAAATCTCGTCGGAAGGACACTCAC   720

Qy  2541  --GCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATACCGTCTCCTTGAAGAGGA  2598
            |||||| ||||||||||||||||||||| |||||||| |||| ||||||||||||||||
Db   721  CGGCAGTTTCTGGATCACCTGTGTATGTCACCAGAGAGGGATACAGTCAGTCTCCTTGAAGAGGA   780

Qy  2599  A  2599
          |
Db   781  A   781
```

RESULT 13
CD243351
LOCUS   CD243351   940 bp   mRNA   linear   EST 22-MAY-2003
DEFINITION AGENCOURT_14121251 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30383283 5', mRNA sequence.
ACCESSION   CD243351
VERSION   CD243351.1 GI:31003815
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 940)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM446 row: o column: 04
            High quality sequence start: 18
            High quality sequence stop: 695.
FEATURES   Location/Qualifiers
     source   1..940
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30383283"
                     /lab_host="DH10B-TonA (T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site 2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

Query Match   22.0%; Score 751.4; DB 14; Length 940;
Best Local Similarity   99.5%; Pred. No. 1.4e-136;
Matches 785; Conservative   0; Mismatches   1; Indels   1; Gaps   3;

```
Qy   125  TCCGGCAGCTCTTTGGGGACAAGGAAGCTGCCCTGCGCGGGACAGGTCCTGCTGAAGTCCG   184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    28  TCCGGCAGCTCTTTGGGGACAAGGAAGCTGCCCTGCGCGGGACAGGTCCTGCTGAAGTCCG    87

Qy   185  GCTACGCGTTCGTGGACTACCCGACCAGAACTGGGCCATCCGCGCCATCGAGACCCCTCT   244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    88  GCTACGCGTTCGTGGACTACCCGACCAGAACTGGGCCATCCGCGCCATCGAGACCCCTCT   147

Qy   245  CGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGC   304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   148  CGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGC   207
```

```
Qy   305  TAAGGAGCAGGAAATTCAGATTGCAAACATCCTCCTCACCTGCAGTGGGAGGTGTTGG   364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   208  TAAGGAGCAGGAAATTCAGATTGCAAACATCCTCCTCACCTGCAGTGGGAGGTGTTGG   267

Qy   365  ATGGACTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGACACACAG   424
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   268  ATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACAGACACACAG   327

Qy   425  AAACCGGCCGTGTCAACGTCACATATGCACAGAGAGAAGCAAAAATAGCCATGGAGA   484
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   328  AAACCGCCGTGTCAACGTCACATATGCACAGAGAGAAGCAAAAATAGCCATGGAGA   387

Qy   485  AGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAGATTTCCTACATCCCGGATGAAG   544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   388  AGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAGATTTCCTACATCCCGGATGAAG   447

Qy   545  AGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGAGC   604
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   448  AGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGAGC   507

Qy   605  AAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCC   664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   508  AAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGGATCC   567

Qy   665  TGGTCCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGAACA   724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   568  TGGTCCCCACCCAGTTGTTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGAACA   627

Qy   725  TCACTAAGCAGACCAGTCCCGGGTAGAATATCCATAGAAAAAGAGAACTCTGGAGCTGCAG   784
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   628  TCACTAAGCAGACCAGTCCCGGGTAGATATCCATAGAAAAAGAGAACTCTGGAGCTGCAG   687

Qy   785  AGAGGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTC   844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   688  AGAGGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTC   747

Qy   845  TTGAAATCATGCA-GAAAGAGGCAGGATGAGACCAAAACTAG-CCGAAGAGATT-CCTCTGA   901
          |||||||||||| |||||||||||||||||||||||||||| |||||||||||| ||||||
Db   748  TTGAAATCATGCAGGAAGAGGCAGGATGAGACCAAACTAGCCCGAAGAGATTCCTCTGA   807

Qy   902  AAATCTTGG   910
          |||||||||
Db   808  AAATCTTGG   816
```

RESULT 14
CF593505
LOCUS   CF593505   824 bp   mRNA   linear   EST 26-SEP-2003
DEFINITION AGENCOURT_15622249 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30531076 5', mRNA sequence.
ACCESSION   CF593505
VERSION   CF593505.1 GI:36347108
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 824)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:

```
                                                    Db   698 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCANTAAG 757
                                                              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                    Qy   721 AACATCACTAAGCAGAGACCCAGTCC-GGTAGTATCCATA-GAAAAGAGAACTCTG 776
                                                              ||||||||||||||||||||||||| ||||||||||||| |||||||||||||
                                                    Db   758 AACATCACTAAGCAGAGACCCAGTCCCGGGGTAGATATCCTTAGGAAAAGAGAACTCTGG 815
```

RESULT 15
BG748346
LOCUS   BG748346   889 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602705902F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842553 5',
            mRNA sequence.
ACCESSION   BG748346
VERSION   BG748346.1   GI:11408999
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1677 row: c column: 02
            High quality sequence stop: 773.
FEATURES   Location/Qualifiers
     source   1..889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842553"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH MGC 43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
ORIGIN

Query Match   21.9%; Score 746; DB 12; Length 889;
Best Local Similarity   95.5%; Pred. No. 1.6e-135;
Matches 802; Conservative   0; Mismatches 30; Indels 8; Gaps 3;

Qy   72 GATGAACAAGCTTACATCGGAACCTGAGCCCCGCCGTCACCGCGACGACCTCCGGCA 131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 GATGAACAAGCTTTCATCGGAACCTGAGCCCCGCCGTCACCGCGACGACCTCCGGCA 61

Qy 132 GCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACAGGTCCTGCTGAAGTCCTACGC 191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GCTCTTTGGGGACAGGAAGCTGCCCCTGGCGCGGGACAGGTCCTGCTGAAGTCCTACGC 121

Qy 192 CTTCGTGGACTACCCCGACCAGAGACTGGGACATCCGGCCATCGAGACCCTCTCGGGTA 251
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 122 CTTCGTGGACTACCCCGACCAGAGACTGGGGCCATCCGCCATCGAGACCCTCTCGGGTAA 181

Qy 252 AGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAG 311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182 AGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAG 241
```

```
                                http://image.llnl.gov
                                Plate: NDAM621 row: e  column: 05
                                High quality sequence stop: 682.
                     FEATURES             Location/Qualifiers
                          source          1..824
                                          /organism="Homo sapiens"
                                          /mol_type="mRNA"
                                          /db_xref="taxon:9606"
                                          /clone="IMAGE:30531076"
                                          /tissue_type="Human Placenta"
                                          /lab_host="DH10B TonA"
                                          /clone_lib="NIH MGC 147"
                                          /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                                          alI-XhoI; Site_2: BamH; Oligo-dT primed using primer
                                          5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                                          insert size 2.3 kb and normalized to ROT 5. This is a
                                          primary library enriched for full-length clones and
                                          constructed using the Cap-trapper method (Carninci, in
                                          preparation). Library constructed by M. Brownstein
                                          (NIMH/NHGRI, National Institutes of Health). Note: This is
                                          a NIH_MGC library."
                     ORIGIN

                     Query Match             22.0%;  Score 749.8;  DB 14;  Length 824;
                     Best Local Similarity   99.4%;  Pred. No. 2.9e-136;
                     Matches 773;  Conservative   0; Mismatches   3;  Indels   2;  Gaps  2;

                     Qy    1 GGCAGGGAGGAGGCGAGGAGGCGCCGGGTACCGGCGGGGGAGCCGGCGGCTCCGGGG 60
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db   38 GGCAGGGAGGAGGCGAGGAGGCGCCGGGTACCGGCGGGGGAGCCGGCGGCTCCGGGG 97

                     Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAAACTGAGCCCCGCCGTCACCGCCGAC 120
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db   98 AAGAGACGGATGATGAACAAGCTTTACATCGGGAAACTGAGCCCCGCCGTCACCGCCGAC 157

                     Qy  121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAG 180
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  158 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAG 217

                     Qy  181 TCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCCGGCCATCGAGACC 240
                              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  218 TCCGGCTACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCCGCGAGACC 277

                     Qy  241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  278 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 337

                     Qy  301 AAGCTAAGGAGCAGGAAAATTCAGATTGGAAACTACCCTCCCCACCTGCAGTGGGAGGTG 360
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  338 AAGCTAAGGAGCAGGAAAATTCAGATTGGAAACTACCCTCCCCACCTGCAGTGGGAGGTG 397

                     Qy  361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 420
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  398 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC 457

                     Qy  421 ACAGAAACCGCCGTTGTCAAGCTGTCACATGTCACATATGCAACAAGAGAAGAAAATAGCCATG 480
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  458 ACAGAAACCGCCGTTGTCAAGCTGTCACATATGCAACAAGAGAAGAAAATAGCCATG 517

                     Qy  481 GAGAAGCTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  518 GAGAAGCTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 577

                     Qy  541 GAAGAGGTGAGCTCCCCTTCGCCCCTTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGG 600
                              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  578 GAAGAGGTGAGCTCCCCTTCGCCCCTTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGG 637

                     Qy  601 GAGCAAGGCCACGCCGCCCCTGGGGGACACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGG 660
                              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                     Db  638 GAGCAAGGCCACGCCGCCCCTGGGGGACACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGG 697

                     Qy  661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720
                              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Thu Jul 22 08:21:46 2004   us-09-270-437d-6.rst

```
Qy    312  CAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGACT  371
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242  CAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGACT  301

Qy    372  TTTGGCTCAATATGGGACAGTGGAACAGTCAACACAGAGAAACCGC  431
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    302  TTTGGCTCAATATGGGACAGTGGAACAGTCAACACAGAGAAACCGC  361

Qy    432  CGTTGTCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGAGAAGCTAAG  491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362  CGTTGTCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGAGAAGCTAAG  421

Qy    492  CGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAG  551
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422  CGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAG  481

Qy    552  CTCCCCTTCGCCCCCTCAGCGAGCCAGCGGGGACCACTCTTCCCGGGAGCAAGGCCA  611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482  CTCCCCTTCGCCCCCTCAGCGAGCCAGCGGGGACCACTCTTCCCGGGAGCAAGGCCA  541

Qy    612  CGCCCCTGGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCCGATCCTGGTCCC  671
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542  CGCCCCTGGGGGCACTTCTCAGCCCAGACAGATTGATTTCCCGCTGCCGATCCTGGTCCC  601

Qy    672  CACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAGAACATCACTAA  731
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602  CAACCAGTTTGTAGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAGAACATCACTAA  661

Qy    732  GCAGACCCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAGCC  791
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    662  GCAGACCCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGAAG-C  720

Qy    792  TGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAA--GCATGCCGCATGCATGATTCTTGAA  849
           ||||||||||||||||||||||||||||||||||||    |||||||||||||||||||
Db    721  TGTCACCATCCATGCCATGCCACCCCAGAGGGGACTTCTTAAAGCATTGCCGCATGATTCTCTGA  780

Qy    850  ATCATGCAGAAAGAGGC-----AGATGACACCAAACTAGCCGAAGAGATTCCTCTGAAAA  904
           |||||||          | |||| || ||| |    ||| ||| |
Db    781  AATCCTGGCAGACAGACCCGCAGATTTGAGACCAAAATAGCCGAAGAAATCGTCTGTAAA  840
```

Search completed: July 21, 2004, 06:03:31
Job time : 8283 secs

```
OM nucleic - nucleic search, using sw model

           GenCore version 5.1.6
     Copyright (c) 1993 - 2004  Compugen Ltd.

Run on:          July 20, 2004, 21:15:35 ; Search time 12992 Seconds
                                            (without alignments)
                                11382.884 Million cell updates/sec

Title:           US-09-270-437D-6
Perfect score:   3412
Sequence:        1 ggcacgcgaggaggcgagga.........aaccttgaaaatgttattt 3412

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters:    6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :       GenEmbl:*
            1:  gb_ba:*
            2:  gb_htg:*
            3:  gb_in:*
            4:  gb_om:*
            5:  gb_ov:*
            6:  gb_pat:*
            7:  gb_ph:*
            8:  gb_pl:*
            9:  gb_pr:*
           10:  gb_ro:*
           11:  gb_sts:*
           12:  gb_sy:*
           13:  gb_un:*
           14:  gb_vi:*
           15:  em_ba:*
           16:  em_fun:*
           17:  em_hum:*
           18:  em_in:*
           19:  em_mu:*
           20:  em_om:*
           21:  em_or:*
           22:  em_ov:*
           23:  em_pat:*
           24:  em_ph:*
           25:  em_pl:*
           26:  em_ro:*
           27:  em_sts:*
           28:  em_un:*
           29:  em_vi:*
           30:  em_htg_hum:*
           31:  em_htg_inv:*
           32:  em_htg_other:*
           33:  em_htg_mus:*
           34:  em_htg_pln:*
           35:  em_htg_rod:*
           36:  em_htg_man:*
           37:  em_htg_vrt:*
           38:  em_sy:*
           39:  em_htgo_hum:*
           40:  em_htgo_mus:*
           41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3411	100.0	3412	6	AR171865	AR171865 Sequence
2	3411	100.0	3412	6	AR343074	AR343074 Sequence
3	3411	100.0	3412	6	BD209925	BD209925 Isolated
4	3383	99.2	3633	9	BC021290	BC021290 Homo sapi
5	3143	92.1	3283	6	AR171867	AR171867 Sequence
6	3143	92.1	3283	6	AR343076	AR343076 Sequence
7	3143	92.1	3283	6	BD209927	BD209927 Isolated
8	3114.8	91.3	3667	9	AF057352	AF057352 Homo sapi
9	2662	78.0	142971	9	AC020629	AC020629 Homo sapi
10	2644	77.5	91084	9	AL596177	AL596177 Human DNA
11	2642.4	77.4	182695	9	AC015706	AC015706 Homo sapi
12	2189.6	64.2	187226	9	AC104980	AC104980 Homo sapi
13	2188	64.1	113201	9	AP004290	AP004290 Homo sapi
14	2165	63.5	3557	10	BC023758	BC023758 Mus muscu
15	1979.8	58.0	2010	9	AF117107	AF117107 Homo sapi
16	1627	47.7	98945	9	AC016961	AC016961 Homo sapi
17	1618	47.4	2640	9	HSM806243	BXS37583 Homo sapi
18	1613	47.3	208273	2	AC134942	AC134942 Homo sapi
19	1234.4	36.2	4953	10	BC054552	BC054552 Mus muscu
20	743.8	21.9	156398	2	AC094924	AC094924 Rattus no
21	743.8	21.8	2021	5	AF026527	AF026527 Gallus ga
22	715.4	21.0	2105	10	AF541940	AF541940 Rattus no
23	712	20.9	1740	6	BD275902	BD275902 COMPOUNDS
24	712	20.9	1740	6	AR220687	AR220687 Sequence
25	712	20.9	1740	6	AR281251	AR281251 Sequence
26	712	20.9	1740	6	AX365954	AX365954 Sequence
27	712	20.9	1743	6	AX366054	AX366054 Sequence
28	712	20.9	1743	6	AX366057	AX366057 Sequence
29	710.4	20.8	4155	9	HSU76705	U76705 Human putat
30	710.4	20.8	4159	6	AR171863	AR171863 Sequence
31	710.4	20.8	4159	6	AR343072	AR343072 Sequence
32	710.4	20.8	4159	6	BD209923	BD209923 Isolated
33	710.4	20.8	4171	9	AF117108	AF117108 Homo sapi
34	710.4	20.8	4181	6	BD275767	BD275767 COMPOUNDS
35	710.4	20.8	4181	6	AR220552	AR220552 Sequence
36	710.4	20.8	4181	6	AR255546	AR255546 Sequence
37	710.4	20.8	4181	6	AR281116	AR281116 Sequence
38	710.4	20.8	4181	6	AX333233	AX333233 Sequence
39	710.4	20.8	4181	6	AX365782	AX365782 Sequence
40	710.4	20.8	4181	9	HSU97188	U97188 Homo sapien
41	710.4	20.8	4602	6	AX397963	AX397963 Sequence
42	709.8	20.8	2697	5	AF064633	AF064633 Xenopus l
43	709.8	20.8	3147	5	AF042353	AF042353 Xenopus l
44	709.4	20.8	2031	5	BC057700	BC057700 Xenopus l
45	709	20.8	1764	6	AX366035	AX366035 Sequence

                              ALIGNMENTS

```
RESULT 1
AR171865
LOCUS       AR171865                3412 bp    DNA        linear    PAT 17-DEC-2001
DEFINITION  Sequence 6 from patent US 6297364.
ACCESSION   AR171865
VERSION     AR171865.1  GI:17910815
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3412)
AUTHORS     Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
TITLE       Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
```

```
JOURNAL   Patent: US 6297364-A 6 02-OCT-2001;
FEATURES
     source        1..3412
                   /organism="unknown"
                   /mol_type="unassigned DNA"
ORIGIN

Query Match           100.0%;   Score 3411;   DB 6;   Length 3412;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 3412;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 GGCAGCGGAGGAGGAGGCGGAGGAGGCGCCGGGGTACCGGGGGAGCCGGGGCTCCGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAGCGGAGGAGGAGGCGGAGGAGGCGCCGGGGTACCGGGGGAGCCGGGGCTCCGGGG 60

Qy     61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCGTCACCGCCGAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCGTCACCGCCCGAC 120

Qy    121 GACCTCCGGCAGCTCTTTGGGACAGGAGCTGCCCGTCGGCGGGACAGTCCTGTGAAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCTCCGGCAGCTCTTTGGGACAGGAGCTGCCCGTCGGCGGGACAGTCCTGTGAAG 180

Qy    181 TCCGGCTACGGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCCGGCGCCAGACC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCGGCTACGGCCTTCGTGACTACCCCGACCAGAACTGGGCCATCCGGCGCCAGACC 240

Qy    241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300

Qy    301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCTCCTCCACCTGCAGTGGGAGGTG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCTAAGGAGCACGGAAAATTCAGATTCGAAACATCCTCCTCCACCTGCAGTGGGAGGTG 360

Qy    361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAAGCAAGTCAACACAGAC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAAGCAAGTCAACACAGAC 420

Qy    421 ACAGAAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGCAAAAAATAGCCATG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACAGAAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGCAAAAAATAGCCATG 480

Qy    481 GAGGAAAACTTCTTTAACCCTTCAGAGTTGAGATTCCTACATCCCCGAT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGGAGAGTGAGCTAAGCGGCATCAGTTCAGTTGAGATTCCTACATCCCCGAT 540

Qy    541 GAGGAGCAAGCCACCGCCCCCTGGGGGCACTTCTCAGGCCAGCAGACAGATTGCCGTGCG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGGAGCAAGCCACCGCCCCCTGGGGGCACTTCTCAGGCCAGCAGACAGATTGCCGTGCG 600

Qy    601 GAGCAAGCCACCGCCCCCTGGGGGCACTTCTCAGGCCAGCAGACAGATTGATTTCCCGCTGCGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGCAAGCCACCGCCCCCTGGGGGCACTTCTCAGGCCAGCAGACAGATTGATTTCCCGCTGCGG 660

Qy    661 ATCCTGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAG 720

Qy    721 AACATCACTAAGCAGAGAAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGAGAAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT 780

Qy    781 GCAGAGAGCCTGTCACCATCCATGCCACCCCAGGGGGACTTCTGAAGCATGCCCCATG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGAGAGCCTGTCACCATCCATGCCACCCCAGGGGGACTTCTGAAGCATGCCCCATG 840

Qy    841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGAAAGAGAATTCCTCTG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGAAAGAGAATTCCTCTG 900

Qy    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTG 960

Qy    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy   1021 ATATACACCCGAAAGAACCATCACTGTGAGGGACACAGTGAGGGCCTGTCCCAGTGCT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACACCCGAAAGAACCATCACTGTGAGGGACACAGTGAGGGCCTGTCCCAGTGCT 1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140

Qy   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy   1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGCGGAGCTCCCCCCGTCCCCTACCAC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGCGGAGCTCCCCCCGTCCCCTACCAC 1260

Qy   1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCAGCAGGAGATGTGAATCTCTTCATCCCAACCCAG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCAGCAGGAGATGTGAATCTCTTCATCCCAACCCAG 1320

Qy   1321 TTCCCCGATCATCACTCTTATCCGAGACGGGGGGCACACATCAAACAGCTGGGCAGATTCGCC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCCGATCATCACTCTTATCCGAGACGGGGGGCACACATCAAACAGCTGGGCAGATTCGCC 1380

Qy   1381 GCTGTGGGGCGCCATCATCGGGAAGAGAGCCCAGACGTCAGCGAAGGATGGTCATC 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGGGGCGCCATCATCGGGAAGAGAGCCCAGACGTCAGCGAAGGATGGTCATC 1440

Qy   1441 GGAGCCTCTATCGAAGATTGCCCCTGCGGAAGGCCCAGGGCCCAGTTCAAGGGACGGATCTTTGGGAAACTGAAA 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGCCTCTATCGAAGATTGCCCCTGCGGAAGGCCCAGGGCCCAGTTCAAGGGACGGATCTTTGGGAAACTGAAA 1500

Qy   1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAA 1560

Qy   1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGCAAAGGTGCCTGAAGTCAGTGCCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGCAAAGGTGCCTGAAGTCAGTGCCC 1620

Qy   1621 TCTTCCACAGCTGGCGCGGGGTGATTGGCGGGGTGATTGGCAAAGGTGCAAGACCGTGAACGAACTGACAGCAGAAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGCTGGCGCGGGGTGATTGGCGGGGTGATTGGCAAAGGTGCAAGACCGTGAACGAACTGACAGCAGAAC 1680

Qy   1681 TTAACCAGTGCGAGAAGTCATCGTGCCTCGTGACCAGAAATGAGAAAGTG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACCAGTGCGAGAAGTCATCGTGCCTCGTGACCAGAAATGAGAAAGTG 1740

Qy   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAGATCAGGGAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAGATCAGGGAA 1800

Qy   1801 ATTGTACAACAGGTGAAGCACAGCAGGAGAAATACCCTCAGGAGTCGGCTCACACGCGC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGGTGAAGCACAGCAGGAGAAATACCCTCAGGAGTCGGCTCACACGCGC 1860

Qy   1861 AGCAAGTGAGGCTCCCACAGGCACCAGCACCAGGAGTGAATTGTAGCCCTTCCAAACAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCTCCCACAGGCACCAGCACCAGGAGTGAATTGTAGCCCTTCCAAACAC 1920

Qy   1921 CTGACGAATGAGACAGAACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACGAATGAGACAGAACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAG 1980

Qy   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGACTCGCCGAGGACCCGGAGAACCCCAGGGGC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGACTCGCCGAGGACCCGGAGAACCCCAGGGGC 2040
```

```
Qy  3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGCTACCTCAG 3180

Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240

Qy  3241 TTGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT 3360

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2

AR343074

```
LOCUS       AR343074                3412 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 6 from patent US 6576756.
ACCESSION   AR343074
VERSION     AR343074.1  GI:33738476
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3412)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6576756-A 6 10-JUN-2003;
FEATURES             Location/Qualifiers
     source          1..3412
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

Query Match   100.0%; Score 3411; DB 6; Length 3412;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 3412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GGCAGCGGGAGGAGGCGGAGGAGGCGGCGGAGGCGGGGGAGGAGGCGGAGGCGGAGGCGGAGGCGGAG  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCGGGAGGAGGCGGAGGAGGCGGCGGAGGCGGGGTACCGGGCGGGGAGGAGGCGGAGGCGGCTCGGGG  60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACTTACATCGGAACTTGAGCGCCCGCGTCACCGGCCGAC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACTTACATCGGAACTTGAGCCCCGCGTCACCGGCCGAC 120

Qy  121 GACCTTCCGGCAGCTCTTTGGGGACAAGGAGCTGCCCTGCGGGGACAGGAGTCCTGCTGAAG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GACCTTCCGGCAGCTCTTTGGGGACAAGGAGCTGCCCCTGCGGGGACAGGAGTCCTGCTGAAG 180

Qy  181 TCCGGCTACGCCTTCGTCGTGGACTACCCCGACCTACCCCGACCTCAGTTGATTACTCAGTCTCTAA 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCTTCGTCGTGGACTACCCCGACTACCCCGACCTCAGTTGATTACTCAGTCTCTAA 240

Qy  241 CTCTCGGGTAAAGTCGGAATTCGCATGGGAAATTCAGATTCAGATTCGAAACTTCAGATTCTCAGTCTCTAA 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGGTAAAGTCGGAATTCGCATGGGAAATTGCATGGGAAATTCAGATTCGAAACTTCAGATTCTCAAA 300

Qy  301 AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAGAAATTAGAAATTTACCCAAGGTTGTTGGCC 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGCTAAGGAGCAGGAGAAATTCAGATTCAATATCAGATTCGAAGAAATTTACCCAAGGTTGTTGGCC 360

Qy  361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGAC 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGAC 420
```

```
Qy 2041 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCCAGACCACCGAGCCCGCCTCCCGCC 2100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 CGAGGAGGGGCGGGGAAGTCAGCCAGGTTGCCAGACCACCGAGCCCGCCTCCCGCC 2100

Qy 2101 CCCAGGGCTTCGCAGGCTTCAGCCATCCACTCCACTCCACTCGGATCTCTCCTGAA 2160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 CCCAGGGCTTCGCAGGCTTCAGCCATCCACTCCACTCCACTCGGATCTCTCCTGAA 2160

Qy 2161 CTCCCACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 CTCCCACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy 2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAA 2280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAA 2280

Qy 2281 GGGAAGATGTTAAGTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 GGGAAGATGTTAAGTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2340

Qy 2341 TTTAGAAATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 TTTAGAAATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy 2401 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy 2461 GTCCTCACGGTGTAGAGAGGAGCTTTGAGGCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 GTCCTCACGGTGTAGAGAGGAGCTTTGAGGCACCCGCACAAAATTCACCCAGAGGGAAAT 2520

Qy 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATA 2580
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATA 2580

Qy 2581 CCGTCTCCTTGAAGAGGAAAACTCTGTCACTCGTAGCTCGTCGTAGCTCATACACCCATT 2640
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 CCGTCTCCTTGAAGAGGAAAACTCTGTCACTCGTAGCTCGTCGTAGCTCATACACCCATT 2640

Qy 2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGTTTTTGCATATCGTTTTTGCATATCTTCTCTGTCT 2700
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGTTTTTGCATATCGTTTTTGCATATCTTCTCTGTCT 2700

Qy 2701 CTCTCGTGTTTATTCTCTCCATCTCCCCTCCCCTTCTCCCCATCTCCCATTCCCATTCTTTTGAA 2760
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 CTCTCGTGTTTATTCTCTCCATCTCCCCTCCCCTTCTCCCCATCTCCCATTCCCATTCTTTTGAA 2760

Qy 2761 TTTTCCTCATCCCTCCATCTCCAATCCCGTATCTACGGCACCCCCCCCCCCCCAGGCAAAGCA 2820
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 TTTTCCTCATCCCTCCATCTCCAATCCCGTATCCGTATCTACGGCACCCCCCCCCCCCCAGGCAAAGCA 2820

Qy 2821 GTGCTCTGAGTATCACATCACAACAAAGGCAAAACACACAACCAGCTCAA 2880
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 GTGCTCTGAGTATCACATCACAACAAAGGAACAAAAGCAAAACACACAACCAGCTCAA 2880

Qy 2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGGTTTTTGGAAG 2940
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGGTTTTTGGAAG 2940

Qy 2941 AGGAAAACAGGAACCAACAAACCACCAACCAATCAACCAAACAAAGAAAAATTCCACAATG 3000
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 AGGAAAACAGGAACCAACAAACCACCAACCAATCAACCAAACAAAGAAAAATTCCACAATG 3000

Qy 3001 AAAGAATGTATTTGTCTTTTGTCTTTTGCATTTGGTATAAGCCATCAATATTCAGCAAAATG 3060
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 AAAGAATGTATTTGTATTTTGTCTTTTGCATTTGGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy 3061 ATTCCTTTTCTTTAAAAAAATGTGGAGGAAAATGTGAGGAAGAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 ATTCCTTTTCTTTAAAAAAATGTGGAGGAAAATGTGAGGAAGAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
```

```
Db   361 TTGGATGGACTTTGGCTCAATATGGACAGTGGACGAATGTGGAACAAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTGTCAACGTCACATATGCAACAGAAGAAGCAAAAATAGCCATG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTGTCACGTCACATATGCAACAGAAGAAGCAAAAATAGCCATG 480

Qy   481 GAGAAGTAAGGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT 540

Qy   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCG 600

Qy   601 GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTCCCGCTGCGG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGCCCAGACAGATTGATTCCCGCTGCGG 660

Qy   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAAGAACTCTTGGAGCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCATCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840

Qy   841 ATTCTTGAAATCATGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGATAACAATCTCATCTTTGCAGGATTTGAGC 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAGAAGCAGAAATTTG 900

Qy   901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAGAAGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAGAAGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCGTCTCTGTACATGTCGTTGAGC 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCGTCTCTGTACATGTCGTTGAGC 1020

Qy  1021 ATATACAACCCGGAAAGAACCATCCACACTGTGAAGGGCCACAGTTGAGGCCTGTGCCAGTGCT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAACCATCCACACTGTGAAGGGCCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGAAATGATATGCTGGCTGTTAAC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGAAATGATATGCTGGCTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCGGTTGAACCTCAGCGGGCCACTTCTTCAGCCTGATCTTTTCAACAGGA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGCCAATCTGATCCCGGTTGAACCTCAGCCGGCCACTTCTTCAGCCTGATCTTTTCAACAGGA 1200

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGAGGGGCCCCCGTGCCCCTACCAC 1260
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGCTATCTCCACCAGCAGAGGGGCCCCCGTGCCCCTACCAC 1260

Qy  1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCGTGTACCCCGTGCCCATCACCAGTTTGGCCCG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCGTGTACCCCGTGCCCATCACCACCAGTTTGGCCCG 1320

Qy  1321 TTCCCGCATCATCCGGCCATCATCCGGGAAGAAGGGCACACATCAAACACTGGCCGGAGATTCGCC 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCGCATCATCCGGCCATCATCCGGGAAGAAGGGCACACATCAAACACTGGCCGGAGATTCGCC 1380

Qy  1381 GCTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACACAGTCGGCCGGAGATTCGCC 1440
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCTGTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACACAGTCGGCCGGAGATTCGCC 1440

Qy  1441 GGAGCCTTCTATCAAGATTGCCCCTGCGGAGAGGCCCGGAGGGACACTCACGGCAGTTCTGGAAGGATGTCATC 1500
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTTCTATCAAGATTGCCCCCTGCGGAGAGGCCCGGAGGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATA 1500
```

```
Qy  1501 ATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAGGACGGGATCTTTGGGAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAGGACGGGATCTTTGGGAAACTGAAA 1560

Qy  1561 GAGGAAAACTTCTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAGGAAAACTTCTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620

Qy  1621 TCTTCCACAGCTGCCGGGTGATTGGCAAGGTGGCAAGACCGTGACGAACTGCGAGAAC 1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGCTGCCGGGTGATTGGCAAGGTGGCAAGACCGTGACGAACTGCGAGAAC 1680

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCTCGTGCCCAGATGATTGAGGAGAAGTG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCTCGTGCCCAGATGAAATAGAGGAGAAGTG 1740

Qy  1741 ATCGTCAGAATTATCGGGCACTCTTCTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800

Qy  1801 ATTGTACAACAGTGAAGCAGCAGCAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATTGTACAACAGTGAAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860

Qy  1861 AGCAAGTGAGGTCCCACAGCCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAGGTCCCACAGCCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920

Qy  1921 CTGACAGAATGAGACCAAAACGCAGCCAGCATCGGGAGGACTCTGCCGAGCCCCTGAGAAACCCCCAGGGGC 1980
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACGCTGCGCGGAGGACGGCCAGCATCGGGAGGACTCTGCCGAGCCCCTGAGAAACCCCCAGGGGC 1980

Qy  1981 GAATGAGAGTCTGCGGAGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCCAGGGCC 2040
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAGTCTGCGGAGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCACCCAGGGCC 2040

Qy  2041 CGAGGAGGGGGCCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGGGCCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCAGGCTTCAGCGAGCTTCAGCCATCCACTCCACTCGGATCTCTCCTGCT 2160
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGGCTTCTGCGAGCTTCAGCCATCCACTCCACTCGGATCTCTCCTGAA 2160

Qy  2161 CTCCCACGACGCTATCCCTTTGTTAGTTGAACTAACATAGGTGAACTGTTCAAAGCCAAGC 2220
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTGTTAGTTGAACTAACATAGGTGAACTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGTGCAAATCGTCTCTGTACATGTGTTTAATCAATATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTCTGTGTGCAAATGTCTCTGTACATGTGTTTAATCAATATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCAGCGGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATCAAATAATCAAATAATCCAATTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATCAAATAATCAAATAATCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTCTTTTTAAAGAGAAAGCAGGGTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTCTTTTTAAAGAGAAAGCAGGGTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTTCACGGTGTAGAGAGGGAAATTCACCCGAGGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGGGACTTTGAGGGGCACCCGAGGGGAAAT 2520

Qy  2521 CTCGTCGGAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGGGATA 2580
```

JOURNAL

COMMENT

```
          antigens, these antigens and method of using the same
          Patent: JP 2002512049-A 4 23-APR-2002;
          LUDWIG INSTITUTE FOR CANCER RESEARCH
       OS  Homo sapiens (human)
       PN  JP 2002512049-A/4
       PD  23-APR-2002
       PF  16-MAR-1999 JP 2000545030
       PR  17-APR-1998 US   09/061709
       PI  YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
       PI  JAGER,
       PI  ALEXANDER KNUTH,LLOYD J OLD
       PC  C12N15/09,A61K35/12,A61K39/00,A61P35/39,A61P35/00,C07K16/32,
       PC  C12N1/15,
       PC
       PC  C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
       PC  574,
       PC  C12N15/00,C12N5/00
       CC  n=not determined(3372)
       FH  Key             Location/Qualifiers
       FT  source          1..3412
       FT                  /organism='Homo sapiens (human)'.
```

FEATURES   Location/Qualifiers
source   1..3412
```
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
```

ORIGIN

```
Query Match        100.0%;   Score 3411;   DB 6;   Length 3412;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 3412;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;
```

```
Qy    1 GGCAGCGGAGGAGGCGGAGAGCGCCGCGGGTACCCGCCGGGGGAGGCCGGGGCCTCGGGGG  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCGGAGGAGGCGGAGAGCGCCGCGGGTACCCGCCGGGGGAGGCCGGGGCCTCGGGGG  60

Qy   61 AAGAGACGGATGATGAACAAGCTTACATCGGGACTTTACATCGGGACTCGAGCCGTCACCGCCGAC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTACATCGGGACTTTACATCGGGACTCGAGCCGTCACCGCCGAC 120

Qy  121 GACCTCCGGCAGCTCTTTGGGGACCAGGAAGCTGCCCGGGGACAGTCCTGCTGAAG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GACCTCCGGCAGCTCTTTGGGGACCAGGAAGCTGCCCGGGGACAGTCCTGCTGAAG 180

Qy  181 TCCGGCTACGCCTTCGTGGACTACACCCCGACCACCCGGACAACTGGGCCATCCGAGACC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCTTCGTGGACTACACCCCGACCACCCGGACAACTGGGCCATCCGAGACC 240

Qy  241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300

Qy  301 AAGCTAAGGAGCAGGAAAATTCAGATTGAAACATCCCTCCAGAATGTGGAACAAGTCAACAGAC 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGCTAAGGAGCAGGAAAATTCAGATTGAAACATCCCTCCAGAATGTGGAACAAGTCAACAGAC 360

Qy  361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAAGGAAGAGAAGCAAAAATAGCCATG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAAGGAAGAGAAGCAAAAATAGCCATG 420

Qy  421 ACAGAAACCGCCGTTGTCACGTCACGTGTCAACTACTCCTTCAAGATTTCCTACATCCCGGAT 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ACAGAAACCGCCGTTGTCAACGTCACGTGTCAACTACTCCTTCAAGATTTCCTACATCCCGGAT 480

Qy  481 GAGAAGCTAAGCAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGAAGCTAAGCAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540

Qy  541 GAAGAGGTGAGCTCCCCTTGCCCCCTGCCCGAGCCCCAGCGCTGGGGACCACTCTTTCCCGG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAAGAGGTGAGCTCCCCTTGCCCCCTGCCCGAGCCCCAGCGCTGGGGACCACTCTTTCCCGG 600

Qy  601 GAGCAAGGCCACGCCCCTCGGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG 660
```

```
Qy 2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGCGTCAGCTCATACACCCATT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGCGTCAGCTCATACACCCATT 2640

Qy 2641 TCTCTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 TCTCTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700

Qy 2701 CTCTCTGTTTATCTCCCCTCCCCCCCCCCCTTCTTCCATCTCCATTCTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 CTCTCTGTTTATCTCCCCTCCCCCCCCCCCTTCTTCCATCTCCATTCTTTGAA 2760

Qy 2761 TTTCCATCCCTCCATCTCCATCTCCGTATCTACGCACCCCCCCCCCCCCAGGCAAAGCA 2820
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 TTTCCATCCCTCCATCTCCATCTCCGTATCTACGCACCCCCCCCCCCCCAGGCAAAGCA 2820

Qy 2821 GTGCTCTGAGTATCATCACACAAAAGGAACAAAAGCGAAACACAAACCAGCCTCA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 GTGCTCTGAGTATCATCACACAAAAGGAACAAAAGCGAAACACAAACCAGCCTCA 2880

Qy 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940

Qy 2941 AGGAAAACAGGAACCCACCACCAACCAATCAACCAACCAAACAAAGAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 AGGAAAACAGGAACCCACCACCAACCAATCAACCAACCAAACAAAGAAAAATTCCACAATG 3000

Qy 3001 AAAGAAGTGTATTTGTCTTTTGTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 AAAGAAGTGTATTTGTCTTTTGTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy 3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 ATTCCTTTCTTTAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120

Qy 3121 CAGGGCGTTAAATTCACAGATTTTTAACGAGAAAACACACAGAAAAACACACAGAGAGCTACCTCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3121 CAGGGCGTTAAATTCACAGATTTTTAACGAGAAAACACACAGAAAAACACACAGAGAGCTACCTCAG 3180

Qy 3181 GTGTTTTACCTGCACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3181 GTGTTTTTACCTGCACCTTGCTCTGTGTTCCCTTAGAGATTTGTAAAGTGATAG 3240

Qy 3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAATAAGAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3241 TTGGAGCATTTTTTTATTTTTTAATAAAAATGAGTTGGAAAAATAAGAGATATCAACT 3300

Qy 3301 GCCAGCCTGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCGCAACAGCGTGTTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3301 GCCAGCCTGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTCGCAACAGCGTGTTCTTCCGCT 3360

Qy 3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAAATGTTTATTT 3412
         |||||||||||||||||||||||||||||||||||||||||||
Db 3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAAATGTTTATTT 3412
```

RESULT 3
BD209925
LOCUS   BD209925   3412 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209925
VERSION   BD209925.1 GI:33019695
KEYWORDS   JP 2002512049-A/4.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3412)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE   Isolated nucleic acid molecules encoding cancer-associated

```
Db  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740

Qy  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGACTGCACAGCGCAAGATCAGGGAA 1800

Qy  1801 ATTGTACAACAGTGAGCAGCAGGAGCAGAATACCTCAGGGAGTCGCCTCACAGCGC     1860
         |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  1801 ATTGTACAACAGTGAGCAGCAGGAGCAGAATACCCTCAGGGAGTCGCCTCACAGCGC    1860

Qy  1861 AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920

Qy  1921 CTGACAGAATGAGACCAAACGCAGCAGCCAGATCGGAGCAAACCAAAGACCATCTGAG    1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACCAAACGCAGCAGCCAGATCGGAGCAAACCAAAGACCATCTGAG   1980

Qy  1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGACCCCAGGGGC  2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGACCCCAGGGGC 2040

Qy  2041 CGAGGAGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCGCAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTCCTGAA   2160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCGCAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTCCTGAA  2160

Qy  2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACTAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA  2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGTACATATTAGAAA   2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGGGGTAATATAT 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGGGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTTAAAGAAAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTTAAAGAAAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGAGGACACTCACGGCAGTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGGACACTCACGGCAGTTTGAGGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520

Qy  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGGAGAGGATA 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGGAGAGGATA 2580

Qy  2581 CCGTCTCCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2640

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATATTCTCTGTCT 2700

Qy  2701 CTCTCGTTTATCTCTCCCCTCCCCCCTTCTTCCATCTTTTGAA                 2760
         ||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTCTCGTTTATCTCTCCCCTCCCCCCTTCTTCCATCTTTTGAA                2760

Qy  2761 TTTCCTACCCTCCATCTCAATCCCGTATCTCACCGCACCCCCCCCCCCCCCAGGCAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TTTCCTACCCTCCATCTCAATCCCGTATCTCACCGCACCCCCCCCCCCCCCAGGCAAAGCA 2820
```

```
Db   601                                                      ||||||||||||||||||||
         GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCG 660

Qy   661 ATCCTGGTCCCCACCCAGTTTGTTGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTTGTTGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATCATCGGAAAAGAGAACTCTGGAGCT 780
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCAGTCCCGGGTAGATATCCATCATCGGAAAAGAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGACTTTGAAGCATGCCGCATG    840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCCATGCCACCCAGAGGGACTTTGAAGCATGCCGCATG   840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCACACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCACACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy  1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA 1200

Qy  1201 CTGTCCGTGCTATCTCCACCCAGAGGGCCCCGGGGAGTCCCCCGTGCCCCCTACCAC   1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCGTGCTATCTCCACCCAGAGGGCCCCGGGGAGTCCCCCGTGCCCCCTACCAC    1260

Qy  1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGACTCGTACCCCCATCACCAGTTTGGGCCG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGACTCGTACCCCCATCACCAGTTTGGGCCG 1320

Qy  1321 TTCCCGCATCATCACACTCTTATCCAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTCCCGCATCATCACACTCTTATCCAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380

Qy  1381 GCTGTGGGGGCCATCATCGGGAAGAAGGGGCCACACATCAAACAGCTTGGCGAGATTGGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCTGTGGGGGCCATCATCGGGAAGAAGGGGGCCACACATCAAACAGCTCGGCGAGATTGCCC 1440

Qy  1441 GGAGGCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500

Qy  1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACCGGGCACCGGAAGCCCAGTTCAAGGCCCCAGGGACGGATCTTTGGGAAACTGAAA 1560

Qy  1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCC 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCC 1620

Qy  1621 TCTTCCACAGTCTGGCCGGGGTGATTGGCAAGAGTGCAAGACCGTGAACGAACTGCAGAAC 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGTCTGGCCGGGGTGATTGGCAAGAGTGCAAGACCGTGAACGAACTGCAGAAC 1680

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   2821  GTGCTCTGAGTATCACATCACACAAAGGAACACAAAAGGCGAAACACAAAACCAGCCCTCAA  2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  GTGCTCTGAGTATCACATCACACAAAGGAACACAAAAGGCGAAACACAAAACCAGCCCTCAA  2880

Qy   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG   2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG   2940

Qy   2941  AGGAAAACAGGAACCACCAACCAACCAACCAATCAACCAAAGAAAAATTCCACAATG   3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  AGGAAAACAGGAACCACCAACCAACCAACCAACCAATCAACCAAACAAAGAAAAATTCCACAATG   3000

Qy   3001  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTTATAAGCCATCAATATTCAGCAAAATG   3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTTATAAGCCATCAATATTCAGCAAAATG   3060

Qy   3061  ATTCCTTTCTTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC   3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  ATTCCTTTCTTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC   3120

Qy   3121  CAGGGCGTAAATTCACAGATTTTTTTAACGAGAAGAAAACACACAGAAGAAGAGCTACCTCAG   3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  CAGGGCGTAAATTCACAGATTTTTTTAACGAGAAGAAAACACACAGAAGAAGAGCTACCTCAG   3180

Qy   3181  GTGTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTCGATAG   3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  GTGTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTCGATAG   3240

Qy   3241  TTGGAGCATTTTTTTATTTTTTTAATAAAAATGAGTTGAAAAAAATAAGAATATCAACT   3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241  TTGGAGCATTTTTTTATTTTTTTAATAAAAATGAGTTGAAAAAAATAAGAATATCAACT   3300

Qy   3301  GCCAGCCTGGACGAAGCTGACAGTCCAAGTCGCACAGCAGTGTTCTGAATTGTCTTCCGCT   3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  GCCAGCCTGGACGAAGCTGACAGTCCAAGTCGCACAGCAGTGTTCTGAATTGTCTTCCGCT   3360

Qy   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT   3412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT   3412
```

RESULT 4

BC021290

LOCUS   BC021290   3633 bp   mRNA   linear   PRI 04-OCT-2003
DEFINITION Homo sapiens IGF-II mRNA-binding protein 2, mRNA (cDNA clone
            MGC:29539 IMAGE:5090334), complete cds.
ACCESSION   BC021290
VERSION   BC021290.2 GI:3387041
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 3633)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE 22388257
   PUBMED   12477932
REFERENCE   2 (bases 1 to 3633)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   On Aug 19, 2003 this sequence version replaced gi:18204200.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Matrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 39 Row: o Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 27552765.
FEATURES   Location/Qualifiers
     source   1..3633
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:29539 IMAGE:5090334"
                     /tissue_type="Pancreas, epithelioid carcinoma"
                     /clone_lib="NIH_MGC_42"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene   1..3633
                     /gene="IMP-2"
                     /note="synonym: VICK22"
                     /db_xref="LocusID:10644"
     CDS   68..1864
                     /codon_start=1
                     /product="IGF-II mRNA-binding protein 2"
                     /protein_id="AAH21290.1"
                     /db_xref="GI:18204201"
                     /db_xref="LocusID:10644"
                     /translation="MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDY
                     PDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLA
                     QYGTVENVEQVNTDTETAVVNVYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVS
                     SPSPPQRAQRGDHSSREQGHAPGGTSQARQIDFPLRIIVPTQFVGAIIGKEGLTIKNI
                     TKQTQSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMLIEIMQKEADETKLABEIPL
                     KILAHNGLVGRLIGKEGRNLKKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACA
                     SAEIEIMXKLREAFENDMLAVNQQANLIPGLNLSALGIFSTGLSVLSPPAGPRGAPPA
                     APYHPFTHSGYFSSLYPHQGPFPHHSYPEQEIVNLFIPTQAVGAIIGKKGAHIK
                     QLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIFGKLKEENFFNPKEEVK
                     LEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEVIVPRDQTPDENEEVIVRIIGHFFA
                     SQTAQRKIREIVQQVKQQEQKYPQGVASQRSK"
     misc_feature   74..280
                     /note="RRM; Region: RNA recognition motif"
                     /db_xref="CDD:smart00360"
     misc_feature   650..853
```

```
misc_feature    /note="KH; Region: K homology RNA-binding domain"
                /db_xref="CDD:smart00322"
                1340..1555
                /note="KH; Region: K homology RNA-binding domain"
                /db_xref="CDD:smart00322"
ORIGIN

Query Match         99.2%;    Score 3383;  DB 9;  Length 3633;
Best Local Similarity  99.9%;  Pred. No. 0;
Matches 3405;  Conservative  0;  Mismatches  1;  Indels  2;  Gaps  2;
```

```
Qy      6 CGGGAGGAGGCGAGGAGGCGCGGGTACCGGCGCCGGGGGAGCCGGGGGAGCCGCGGGGAGAG 65
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGAGGAGGCGAGGAGGCGCGGGTACCGGCGCCGGGGGAGCCGGGGGAGCCGCGGGGAGAG 60

Qy     66 ACGGATGATGAACAAGCTTTACATCGGGAACAAGCTTTACATCGGGACACCT 125
Db     61 ACGGATGATGAACAAGCTTTACATCGGGAACAAGCTTTACATCGGGACACCT 120

Qy    126 CCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGCGGGACAGGTCCTGCTGAAGTCCGG 185
Db    121 CCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGCGGGACAGGTCCTGCTGAAGTCCGG 180

Qy    186 CTACGCTTCGTGTGGACTACCCCGACCAGGAACTGGGCCATCGCGGCGACCCCTCTC 245
Db    181 CTACGCTTCGTGTGGACTACCCCGACCAGGAACTGGGCCATCGCGGCGACCCCTCTC 240

Qy    246 GGGTAAAGTCGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTTCTAAAAAGCT 305
Db    241 GGGTAAAGTCGGAATTGCATGGGAAAATCATGGGAAGTTGATTACTCAGTTCTAAAAAGCT 300

Qy    306 AAGGAGCAGGAAAATTCAGATTCAGATTCCCTCCTCCAGTGGGAGGTGGA 365
Db    301 AAGGAGCAGGAAAATTCAGATTCAGATTCCCTCCTCCAGTGGGAGGTGGA 360

Qy    366 TGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGA 425
Db    361 TGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGA 420

Qy    426 AACCGCCGTTGTCAACGTCACATATGCAAGGTCCTTTGGACTCCGGGATGAGAA 485
Db    421 AACCGCCGTTGTCAACGTCACATATGCAAGGTCCTTTGGACTCCGGGATGAGAA 480

Qy    486 GCTAAGCGGGCATCAGTTTGAGAACTACTCCCTCCAGCGTGGGGACCTCCCCGGACCA 545
Db    481 GCTAAGCGGGCATCAGTTTGAGAACTACTCCCTCCAGCGTGGGGACCTCCCCGGACA 540

Qy    546 GGTGAGGCTCCCCTTCGCCCCTCCCCAGCGTGGGGCACCTCTTCCCGGGGACA 605
Db    541 GGTGAGGCTCCCCTTCGCCCCTCCCCAGCGTGGGGCACCTCTTCCCGGGGACA 600

Qy    606 AGGGCCACGGCCCCTGGGGGGCACTCTCAGCCAGACGATTGATTTCCCGCTGCGGATCCT 665
Db    601 AGGGCCACGGCCCCTGGGGGGCACTCTCAGCCAGACGATTGATTTCCCGCTGCGGATCCT 660

Qy    666 GGTCCCACCCAGTTGTTGTTGGTGCCATCATCGGAAAGAGGGCTTGACCATAAAGAACAT 725
Db    661 GGTCCCACCCAGTTGTTGTTGGTGCCATCATCGGAAAGAGGGCTTGACCATAAAGAACAT 720

Qy    726 CACTAAGCAGACCAGCCCAGTCCCGGTAGATATCCCCGAGAGGGAATTCCATAGAAGCTGCAGA 785
Db    721 CACTAAGCAGACCAGCCCAGTCCCGGTAGATATCCCCGAGAGGGAATTCCATAGAAGCTGCAGA 780

Qy    786 GAAGCCTGTCACCATCATGCCACCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCCATGATTCT 845
Db    781 GAAGCCTGTCACCATCATGCCACCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCCATGATTCT 840

Qy    846 TGAAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGAGATTCCTCTGAAAAT 905
Db    841 TGAAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGAGATTCCTCTGAAAAT 900

Qy    906 CTTGGCCACCAATGGCTTGGTTGGAAGACTGATTGATTGGAAAGAAGAGGCAGAAATTTGAAGAA 965
```

```
Db    901 CTTGGCACCAATGGCTTGGTTGGAAGACTGATTGATTGGAAAGAAGAAGGCAGAAATTTGAAGAA 960

Db    966 AATTGAACATGAACACAGGAGACCAAGATAACATCTCATCTTTGCAGGATTTGAGCATATA 1025
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AATTGAACATGAACACAGGAGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATA 1020

Qy   1026 CAACCCGGAAAGAACCATCACTGTGAAGGCACAGTGAGGCGTCTGCCAGTGCTGAGAT 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CAACCCGGAAAGAACCATCACTGTGAAGGCACAGTGAGGCGTCTGCCAGTGCTGAGAT 1080

Qy   1086 AGAGATTATGAAGAAGTCGTGTCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACA 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGAGATTATGAAGAAGTCGTGTCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAACCAACA 1140

Qy   1146 AGCCAATCTGATCCCAGGTTGAACCTCAGCGCCACTTGGCATCTTTTCAACAGGACTGTC 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AGCCAATCTGATCCCAGGTTGAACCTCAGCGCCACTTGGCATCTTTTCAACAGGACTGTC 1200

Qy   1206 CGTGCTATCTCCACCAGCAGGGCCCCGGGAGCTCCCCCGTGCCCCTACCACCCCTT 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGTGCTATCTCCACCAGCAGGGCCCCGGGAGCTCCCCCGTGCCCCTACCACCCCTT 1260

Qy   1266 CACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCTGTACCCCATCACCAGTTGGCCCGTTCCC 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCTGTACCCCATCACCAGTTGGCCCGTTCCC 1320

Qy   1326 GCATCATCACTCTTATCCGACGCAGGAGAATTGTGAATCTCTTCATCCCAACCCAGGCTGT 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCATCATCACTCTTATCCGACGCAGGAGAATTGTGAATCTCTTCATCCCAACCCAGGCTGT 1380

Qy   1386 GGGCGCCATCATCGGGAAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCCGGAGC 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GGGCGCCATCATCGGGAAGAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCCGGAGC 1440

Qy   1446 CTCTATCAAGATTGCCCCTGCCGGAAGCGTCAGCGGAAAGGATGGTCATCATCAC 1505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CTCTATCAAGATTGCCCCTGCCGGAAGCGTCAGCGGAAAGGATGGTCATCATCAC 1500

Qy   1506 CGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACCGATCTTGGGAAACTGAAAGAGGA 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACCGATCTTTGGGAAACTGAAAGAGGA 1560

Qy   1566 AAACTTCTTTTAACCCAAAGAAGAAGTGAGCGCATATCAGAGTGCCCTCTTC 1625
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AAACTTCTTTTAACCCAAAGAAGAAGTGAGCGCATATCAGAGTGCCCTCTTC 1620

Qy   1626 CACAGCTGGCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCGAGAACTTAAC 1685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CACAGCTGGCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCGAGAACTTAAC 1680

Qy   1686 CAGTGCAGAGTCATCGTGCCTGTGACCAAAGCCAGATGAAAATGAGGAGAGTGATCGT 1745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGTGCAGAGTCATCGTGCCTGTGACCAAAGCCAGATGAAAATGAGGAGAGTGATCGT 1740

Qy   1746 CAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGT 1805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGT 1800

Qy   1806 ACAACAGGTGAAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACGCCAGCAGCA 1865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACAACAGGTGAAGCAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACGCCAGCAA 1860

Qy   1866 GTGAGGCTCCCACAGCACCACCAGCAGCAAAACACCAGCAGCGGACGATGAATGTAGCCCTTCCAACACCTGAC 1925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTGAGGCTCCCACAGCACCACCAGCAGCAAAACACCAGCAGCGGACGATGAATGTAGCCCTTCCAACACCTGAC 1920

Qy   1926 AGAATGAGAACCAGCAGCCAGCCAGCAAACCAAAGACCATCTGAGGAATG 1985
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AGAATGAGAACCAGCAGCCAGCCAGCAAACCAAAGACCATCTGAGGAATG 1980

Qy   1986 AGAAGTCTGCGGAGGCGGCGGCCTGCGGAGGCCCTGAGAACCCCAGGGGCCGAGG 2045
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AGAAGTCTGCGGAGGCGGCGGCCTGCGGAGGCCCTGAGAACCCCAGGGGCCGAGG 2040
```

```
Qy  3125  GCGTTAAATTCACAGATTTTTTAACGAGAAAAACACAGAGAAGAAGCTACCTCAGGTGT  3184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3120  GCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACAGAGAAGAAGCTACCTCAGGTGT  3179

Qy  3185  TTTTACCCTCAGCACCTTGCTCTTGTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3180  TTTTACCCTCAGCACCTTGCTCTTGTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTTGG  3239

Qy  3245  AGCATTTTTTATTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA  3304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3240  AGCATTTTTTTATTTTTAATAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA  3299

Qy  3305  GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC  3364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3300  GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTAGCC  3359

Qy  3365  AAGAACCNATATGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATT  3412
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  3360  AAGAACCTATATGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATT  3407
```

RESULT 5
```
AR171867
LOCUS       AR171867                3283 bp    DNA        linear    PAT 17-DEC-2001
DEFINITION  Sequence 8 from patent US 6297364.
ACCESSION   AR171867
VERSION     AR171867.1  GI:17910817
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3283)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6297364-A 8 02-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..3283
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

```
Query Match    92.1%;  Score 3143;  DB 6;  Length 3283;
Best Local Similarity   96.2%;  Pred. No. 0;
Matches 3283;  Conservative 0;  Mismatches 0;  Indels 129;  Gaps 1;

Qy    1  GGCAGCGGAGGAGGCGAGGAGCGCCGGGCGCCGGGGAGCCGGCGGGGCTCTCGGGG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GGCAGCGGAGGAGGCGAGGAGCGCCGGGCGCCGGGGAGCCGGCGGGGCTCTCGGGG   60

Qy   61  AAGAGACGGATGATGAACAAGCTTTACATCGGGAAGCTTTACATCGGGAAGCTTT   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AAGAGACGGATGATGAACAAGCTTTACATCGGGAAGCTTTACATCGGGAAGCTTT   120

Qy  121  GACCTCCGCCAGCTCTTTGGGGACGAGGAGCTGCCCTGCGGGACAGGTCCTGCTGAAG  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GACCTCCGCCAGCTCTTTGGGGACGAGGAGCTGCCCTGCGGGACAGGTCCTGCTGAAG  180

Qy  181  TCCGGCTACGCCTTCGTGGACTACCCCGACTACCCCGCCATCCGGCCATCCGAGACC  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCCGGCTACGCCTTCGTGGACTACCCCGACTACCCCGCCATCGGGCCATGCGAGACC  240

Qy  241  CTCTCGGGTAAGTGGAATTGGAATTGCATGGGAAAATCATGGAAATTACTCAGTTCTAAA  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CTCTCGGGTAAGTGGAATTGGAATTGCATGGGAAAATCATGGAAATTACTCAGTTCTAAA  300

Qy  301  AAGCTAAGGAGCAGGAAAAATTCAGATTCGAAACATCCCTCCTCCACCTGCAGTGGAGGTG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AAGCTAAGGAGCAGGAAAAATTCAGATTCGAAACATCCCTCCTCCACCTGCAGTGGAGGTG  360

Qy  361  TTGGATGGACTTTGGCTCAATATGGACAGTGGACAAGTGAACAAGTGCAACACAGAC  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  2046  AGGGGCGGGAAGGTCAGCCAGGTTGCCAGACCACCGAGCCCGCCTTCCCGCCCCCA  2105
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  AGGGGCGGGAAGGTCAGCCAGGTTGCCAGACCACCGAGCCCGCCTTCCCGCCCCCA  2100

Qy  2106  GGGCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCGGATCTCCTGAACTCC  2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  GGGCTTCTGCGAGGCTTCAGCCATCCACTTCACCATCCGGATCTCCTGAACTCC  2160

Qy  2166  ACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTCAAGCCAAGCAAAAT  2225
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  ACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTCAAGCCAAGCAAAAT  2220

Qy  2226  GCACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTCGTACATATTAGAAAGGGAA  2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  GCACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTCGTACATATTAGAAAGGGAA  2280

Qy  2286  GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGGTAATATTTTAG  2345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  GATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGGTAATATTTTAG  2340

Qy  2346  AAATAATATATCAAATAACTCCAATTTTAATCAATTAATCAATTAATTTTTTTT  2405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  AAATAATATATCAAATAACTCCAATTTTAATCAATTAATCAATTAATTTTTTTT  2400

Qy  2406  CTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTAAAGATAAAGTCTTTGGGAGGTCTC  2465
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  CTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCTTTGGGAGGTCTC  2460

Qy  2466  ACGGTGTAGAGGAGAGCTTTGAGGGCCACCCGCACAAAAAATTCACCCGAGAGGGAAATCTCGT  2525
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  ACGGTGTAGAGGAGAGCTTTGAGGGCCACCCGCACAAAAAATTCACCCGAGAGGGAAATCTCGT  2520

Qy  2526  CGGAAGGACACTCACGGCAGTTCTGATCACTGTGTCAACAGAAGGGATACCGTC  2585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CGGAAGGACACTCACGGCAGTTCTGATCACTGTGTCAACAGAAGGGATACCGTC  2580

Qy  2586  TCCTTGAAGAGGAAACTCTGTCACTCCTGTCAGCTCATACACCCATTTCTCT  2645
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TCCTTGAAGAGGAAACTCTGTCACTCCTGTCAGCTCATACACCCATTTCTCT  2640

Qy  2646  TTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTC  2705
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  TTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTC  2700

Qy  2706  TGTTTATCTCTCCCCCTCCCCCTCCCCCTCCCATCTCCATTCTTTTGAATTTCC  2765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  TGTTTATCTCTCCCCCTCCCCCTCCCCCTCCCATCTCCATTCTTTTGAATTTCC  2760

Qy  2766  TCATCCCTCCAATCTCAATCCCGTATCTCACGGCAAAGCAGTGCT  2825
          ||||||||||||||||||||||||||||||||||||||||||||
Db  2761  TCATCCCTCCAATCTCAATCCCGTATCTACGGCA-CCCCAGGCAAAGCAGTGCT  2819

Qy  2826  CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTAC  2885
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2820  CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAACCAGCCTCAACTTAC  2879

Qy  2886  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2945
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2880  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2939

Qy  2946  AACAGGAACCACCAACCAACCAATCAACAACAAAGAAAAATTCCACAATGAAAGA  3005
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2940  AACAGGAACCACCAACCAACCAATCAACAACAAAGAAAAATTCCACAATGAAAGA  2999

Qy  3006  ATGTATTTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC  3065
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3000  ATGTATTTTTGTCTTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCC  3059

Qy  3066  TTTCTTT-AAAAAAAAAAAAAAAAAAGGTTGTTGGGCCCAGG  3124
          ||||||| ||||||||||||||||||||||||||||||||||
Db  3060  TTTCTTTAAAAAAAAAAAAAAAAAAAGGTTGTTGGGCCCAGG  3119
```

Db 361 TTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGACAAGTCAACACAGAC 420

Qy 421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAAATAGCCATG 480
         ||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAAATAGCCATG 480

Qy 481 GAGAAGCTAAGCGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| .
Db 481 GAGAAGCTAAGCGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT 540

Qy 541 GAAGGAGTGAGCTCCCCTTCCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCCG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 GAAGGAGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCCG 600

Qy 601 GAGCAAGGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 GAGCAAGGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 660

Qy 661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG 720

Qy 721 AACATCACTAAGCACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 AACATCACTAAGCACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780

Qy 781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG 840

Qy 841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGCAAAACTAGCCGAAGAGGCAGAAATTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGCAAAACTAGCCGAAGAGGCAGAAATTG 900

Qy 901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAAATTTG 960

Qy 961 AAGAAAATTGAACATGAAACATGAAAGAACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 AAGAAAATTGAACATGAAACATGAAAGAACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy 1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGCCAGTTGAGGCCTGTGCCAGTGCT 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGCCAGTTGAGGCCTGTGCCAGTGCT 1080

Qy 1081 GAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAAC 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTAAC 1140

Qy 1141 CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCCGCCACTTGGCATCTTTTCAACAGGA 1200
Db 1141 -- 1140

Qy 1201 CTGTCCCGTGCTATCTCCACCAGCAGGGCGCCCGCGGAGCTCCCCCGTTGCCCCCCTACCAC 1260
Db 1141 -- 1140

Qy 1261 CCCTTCACTACTCCGGATACTTCTTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG 1320
         |||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 ------ACCCACTCCGGATACTTCTTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG 1191

Qy 1321 TTTCCCGCCATCATCACTCTTATCCAGCAGAGCAAGAAGCAGCTTTCATCCCAACCCAG 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1192 TTTCCCGCCATCATCACTCTTATCCAGCAGAGCAAGAAGCAGCTTTCATCCCAACCCAG 1251

Qy 1381 GCTGTGGGCGCCATCATCGGGGAAGAAGGGGGCACACATCAAACAGCTGGCAGATTCGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1252 GCTGTGGGCGCCATCATCGGGGAAGAAGGGGGCACACATCAAACAGCTGGCAGATTCGCC 1311

Qy 1441 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGCGAAAGATGGTCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1312 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGCGAAAGATGGTCATC 1371

Qy 1501 ATCACCGGCCACCGGAGCCGAAGCCCAGTTCAAGCGCACGGACGGATCTTTGGGAAACTGAAA 1560
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1372 ATCACCGGCCACCGGAGCCCAGTTCAAGCGCACGGACGGATCTCTTGGGAAACTGAAA 1431

Qy 1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGGAGCGCCATATCAGAGTGCCC 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1432 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGGAAGCGCCATATCAGAGTGCCC 1491

Qy 1621 TCTTCCACAGCTGGCCGGTGATTGGCAAGACCGTGGCAAGACTGCAGAAC 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1492 TCTTCCACAGCTGGCCGGTGATTGGCAAGACCGTGGCAAGACTGCAGAAC 1551

Qy 1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCTCGTGACCAAAAGCCAGATGAAAAATGAGGAAGTG 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1552 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGCCTCGTGACCAAAAGCCAGATGAAAAATGAGGAAGTG 1611

Qy 1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGATCGCACAGCGCAAGATCAGGGAA 1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1612 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCAGATCGCACAGCGCAAGATCAGGGAA 1671

Qy 1801 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1672 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1731

Qy 1861 AGCAAGTGAGGGTCCCACAGGCACCACAGCAAAACAACGATGAATGTAGCCCTTCAACAC 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1732 AGCAAGTGAGGGTCCCACAGGCACCACAGCAAAACAACGATGAATGTAGCCCTTCAACAC 1791

Qy 1921 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCGAG 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1792 CTGACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCGAG 1851

Qy 1981 GAATGAGAAGTCTGCGGAGGCGCCAGGAGCTCTGCCAGAACACCAGCCCCGAGGCCTCCCGCGC 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1852 GAATGAGAAGTCTGCGGAGGCGCCAGGAGCTCTGCCAGAACACCAGCCCCGAGGCCTCCCGCC 1911

Qy 2041 CGGAGGAGGGGCGGGAAGGTCAGCGAGGCTCAGCGGTTTGCCAGAACTCACTCCACTCGGATCTCTCCTGAA 2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1912 CGGAGGAGGGGCGGGAAGGTCAGCGAGGCTCAGCGGTTTGCCAGAACTCACTCCACTCGGATCTCTCCTGAA 1971

Qy 2101 CCCCAGGGCTTCTGCAGGCTTCGAGGGTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1972 CCCCAGGGCTTCTGCAGGCTTCGAGGGTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2031

Qy 2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2032 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2091

Qy 2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2280
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 2092 AAAATGCACACCCCTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2151

Qy 2281 GGGAGAGTGTTAAGATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2152 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT 2211

Qy 2341 TTTGAAATAATATCAAATAACTCAATTTTAATCAATTCCAATTTTAATCAATTATTAATTTT 2400
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2212 TTTAGAAATAATATCAAATAACTCAATTTTAATCAATTTTTAATCAATTATTAATTTT 2271

Qy 2401 TTTTCTTTTAAAGAAGAAAGCCTTTCTAGACTTTGAGCGTTTAAAGAATAAGTCTTTGGGAG 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2272 TTTTCTTTTAAAGAAGAAAGCAGGCTTTTGAGCGTTTAAAGAATAAAGTCTTTGGGAG 2331

Qy 2461 GTCTCACGGTGTAGAGAGACACTCACGGCAGTTCTCGGATCACCCTGTATGTCAACGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2332 GTCTCACGGTGTAGAGAGACACTCACGGCAGTTCTCGGATCACCCTGTATGTCAACGAGGGAAAT 2391

Qy 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTCGGATCACCGTGTATGTCAACGAGAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2392 CTCGTCGGAAGGACACTCACGGCAGTTCTCGGATCACCGTGTATGTCAACGAGAGGGATA 2451

```
FEATURES          Location/Qualifiers
    source        1..3283
                  /organism="unknown"
                  /mol_type="genomic DNA"

ORIGIN

Query Match       92.1%;  Score 3143;  DB 6;  Length 3283;
Best Local Similarity  96.2%;  Pred. No. 0;
Matches 3283;  Conservative  0;  Mismatches  0;  Indels 129;  Gaps  1;

Qy    1 GGGAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCGGCGGCGGGAGCCGGCGGCTCGGGG     60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGGAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCGGCGGCGGGAGCCGGCGGCTCGGGG     60

Qy   61 AAGAGACGGATGATGAACAAGCTTACATCGGGAACCTGAGCCCGCCGTCACCGCCGCCAC    120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTACATCGGGAACCTGAGCCCGCCGTCACCGCCGCCAC    120

Qy  121 GACCTCCGGCAGCTCTTCGGGACAGGAAGCTGCCCCTGGCGGGGACAGGTCCTGCTGCTGAAG    180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GACCTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCCTGGCGGGGACAGGTCCTGCTGCTGAAG    180

Qy  181 TCCGGCTACGCCTTCGTCGACTACCCGCCACCGCCAGAACTGGCCATCCCGCGCCATCCGAGACCC    240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCTTCGTCGACTACCCGCCACCGCCAGAACTGGCCATCCCGCGCCATCCGAGACC    240

Qy  241 CTCTCGGGTAAAGTGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA    300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGGTAAAGTGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA    300

Qy  301 AAGTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTGCACCTGCAGTGGGAGTG    360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTGCACCTGCAGTGGGAGTG    360

Qy  361 TTGGATGGACTTTGGCTCAATATGGGACAGTGGAACAAGTCAACACAGAC    420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGGATGGACTTTGGCTCAATATGGGACAGTGGAACAAGTCAACACAGAC    420

Qy  421 ACAGAAACCGCCGTTGTCAAGTCACATATGCAACAGAGAAGCAAAATAGCCATG    480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ACAGAAACCGCCGTTGTCAAGTCACATATGCAACAGAGAAGCAAAATAGCCATG    480

Qy  481 GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCTACATCCCGGAT    540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCTACATCCCGGAT    540

Qy  541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCCAGCGTGGGGGACCACTCTTCCCGG    600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCCAGCGTGGGGGACCACTCTTCCCGG    600

Qy  601 GAGCAAGGCCACCGCCCCTCGGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCGG    660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GAGCAAGGCCACCGCCCCTCGGGGGCACTTCTCAGGCCAGCAGATTGATTTCCCGCTGCG    660

Qy  661 ATCCTGTCCCCACCCAGTTGTTGGTGCCATCATGGAAGATATCCATAGAAAAGAGAACTCTGGAGCT    720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 ATCCTGTCCCCACCCAGTTGTTGGTGCCATCATGGAAGATATCCATAGAAAAGAGAACTCTGGAGCT    720

Qy  721 AACATCACTAAGCACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGAAGCTCTGGAGCT    780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACATCACTAAGCACAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGAAGCTCTGGAGCT    780

Qy  781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCGCGCATG    840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCGCGCATG    840

Qy  841 ATTCTTGAAATCATGCAGAAGAGAGCCAAACTAGCCGAAGAGATTCCTCTG    900
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 ATTCTTGAAATCATGCAGAAGAGAGCCAGAAGAGACCAAACTAGCCGAAGAGATTCCTCTG    900

Qy  901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGAGCAGAAATTTG    960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGAGCAGAAATTTG    960
```

```
Qy 2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCCTCAGCTCATACACCCATT   2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2452 CCGTCTCCTTGAAGAGGAAACTCGTCACTCCTCATGCCTGTCCTCAGCTCATACACCCATT   2511

Qy 2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTATATAATTCTCTGTCT   2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2512 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTATATAATTCTCTGTCT   2571

Qy 2701 CTCTCGTCTGTTATCTCTCCCCTCCCCTCCCCCCCCTTCCCTTTCTCTCCATCTCCATTCTTTGAA   2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2572 CTCTCGTCTGTTATCTCTCCCCTCCCCTCCCCCCCCTTCCCTTTCTCTCCATCTCCATTCTTTGAA   2631

Qy 2761 TTTCCTCATCCCTCCAATCTCAATCCCGTATCTACGGCACCACCGCCAGGCAAAGCA   2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2632 TTTCCTCATCCCTCCAATCTCAATCCCGTATCTACGGCACCACCGCCAGGCAAAGCA   2691

Qy 2821 GTGCTCTGAGTATCACATCACACAAAAGGAACACAAACCAGCCTCAA   2880
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 2692 GTGCTCTGAGTATCACATCACACAAAAGGAACACAAACCAGCCTCAA   2751

Qy 2881 CTTACACTTGGTTACTCAAAAGAACACAGAGTCAATGGTACTTGTCCTGGTTTTGGAAG   2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2752 CTTACACTTGGTTACTCAAAAGAACACAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG   2811

Qy 2941 AGGAAAACAGGAACCCACCACCACCAACCAATCAACCAAACAAGAAAAATTCCACAATG   3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2812 AGGAAAACAGGAACCCACCACCACCAACCAATCAACCAAACAAGAAAAATTCCACAATG   2871

Qy 3001 AAAGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG   3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2872 AAAGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG   2931

Qy 3061 ATTCCTTCTTAAAAAAAAAAATGTGGAGGAAAGTAGGAAAGTTACCAGGTTGTTGGCC   3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2932 ATTCCTTCTTAAAAAAAAAATGTGGAGGAAAGTAGGAAAGTTACCAGGTTGTTGGCC   2991

Qy 3121 CAGGGCGTAAATTCCACAGATTTTTTAACGAGAAAAACACCAGAAGAGGAGCTACCTCAG   3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2992 CAGGGCGTAAATTCCACAGATTTTTTAACGAGAAAAACACCAGAAGAGGAGCTACCTCAG   3051

Qy 3181 GTGTTTTACCTCAGCACCTTGCTCCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG   3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3052 GTGTTTTACCTCAGCACCTTGCTCCTTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG   3111

Qy 3241 TTGGAGCATTTTTTTATTTTTTATTAATAAAATGAGTTGAAAAAAAAATAAGATATCAACT   3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3112 TTGGAGCATTTTTTTATTTTTTATTAATAAAATGAGTTGAAAAAAAAATAAGATATCAACT   3171

Qy 3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCACAGCTGTTCTGAATTGTCTTCCGCT   3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3172 GCCAGCCTGGAGAAGTGACAGTCCAAGTGTGCACAGCTGTTCTGAATTGTCTTCCGCT   3231

Qy 3361 AGCCAAGAACCNATATGGCCTCTTTTGGACAAACCTTGAAAATGTTTATTT   3412
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3232 AGCCAAGAACCNATATGGCCTCTTTTGGACAAACCTTGAAAATGTTTATTT   3283
```

```
RESULT 6
AR343076
LOCUS        AR343076        3283 bp        DNA        linear        PAT 17-AUG-2003
DEFINITION   Sequence 8 from patent US 6576756.
ACCESSION    AR343076
VERSION      AR343076.1  GI:33738478
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1 (bases 1 to 3283)
  AUTHORS    Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
             and Old,L.J.
  TITLE      Isolated nucleic acid molecule encoding cancer associated antigen,
             the antigen itself, and uses thereof
  JOURNAL    Patent: US 6576756-A 8 10-JUN-2003;
```

Left column:

Qy	961	AAGAAAATTGAACACATGAAACACAGGGACCAAGATAACAATTCTCATCTTTGCAGGATTTGAGC 1020
Db	961	AAGAAAATTGAACACATGAAACACAGGGACCAAGATAACAATTCTCATCTTTGCAGGATTTGAGC 1020
Qy	1021	ATATACAACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
Db	1021	ATATACAACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
Qy	1081	GAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGCGTGGTAAC 1140
Db	1081	GAGAATAGAGATTATGAGAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC 1140
Qy	1141	CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGA 1200
Db	1141	--- 1140
Qy	1201	CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCCGCCTGCCCCCTACCAC 1260
Db	1141	--- 1140
Qy	1261	CCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGGCCG 1320
Db	1141	-------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGGCCG 1191
Qy	1321	TTCCCGCCATCACTCTTATCCGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
Db	1192	TTCCCGCCATCACTCTTATCCGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1251
Qy	1381	GCTGTGTGGGCGGCCATCATCGGGGAAGAAGAGGGGGCACACATCAAAACAGCTCGGCGAGATTCGCC 1440
Db	1252	GCTGTGTGGGCGGCCATCATCGGGGAAGAAGAGGGGGCACACATCAAAACAGCTCGGCGAGATTCGCC 1311
Qy	1441	GGAGCCTCTATCAAGATTGCCCCTGGCAAGAAGCTGAAGCGTCAGGCCCAGAGCGTCATCATC 1500
Db	1312	GGAGCCTCTATCAAGATTGCCCCTGGCAAGAAGCTGAAGCGTCAGGCCCAGAGATGCGTCATC 1371
Qy	1501	ATCACCGGGGCACCCGAAGCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAA 1560
Db	1372	ATCACCGGGGCACCCGAAGCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAACTGAAA 1431
Qy	1561	GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAGCGCATATCAGAGTGCCC 1620
Db	1432	GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAGCGCATATCAGAGTGCCC 1491
Qy	1621	TCTTCCACAGCTGGCGCGGGGTGATTGGCAAAGGTGCAAGACCGTGCAAGCGAACTGCAGAAC 1680
Db	1492	TCTTCCACAGCTGGCGCGGGGTGATTGGCAAAGGTGCAAGACCGTGCAAGCGAACTGCAGAAC 1551
Qy	1681	TTAACCAGTGCAGAAGTGCAGAAGCAGAGCAGCAAAACGCAGAGTGAAAATGAGGAAGTG 1740
Db	1552	TTAACCAGTGCAGAAGTGCAGAAGCAGAGCAGCAAAACGCAGAGTGAAAATGAGGAAGTG 1611
Qy	1741	ATCGTCAGAATTATCGGGGCACTTCTTTGCTGCGTGCACTGCACAGCCAGTCGCACAGCGAC 1800
Db	1612	ATCGTCAGAATTATCGGGGCACTTCTTTGCTGCGTGCACTGCACAGCCAGTCGCACAGCGAC 1671
Qy	1801	AGCAAGTGAGACACAGAGGCCTCCACAGGCACCAGCCAAACAAACGCCTCCAACGC 1860
Db	1672	AGCAAGTGAGACAGAGGCCTCCACAGGCACCAGCCAAACAAACGCCTCCAACGC 1731
Qy	1861	ATTGTACACACAGGTGAAGCAGCAGAATGTAGCCCTTCAGCCGGATGAATGGCGCGC 1920
Db	1732	AGCAGTGAGGAGCTCCCACAGCCAAATGAATGTAGCCCTTCAGCCGGATGAATGGCGCGC 1791
Qy	1921	CTGACAGAATGAGACACAAGCCCAGCGCAAACCAAAGACCATCTGAG 1980
Db	1792	CTGACAGAATGAGACATGAGAAGCCCAGCGCAAACCAAAGACCATCTGAG 1851
Qy	1981	GAATGAGAAGAGTCTGCGGAGGCGGCCCAGGGACTCTGCCCGAGACCCCAGGGGC 2040
Db	1852	GAATGAGAGAGTCTGCGGAGGCGGCCCAGGGACTCTGCCCGAGACCCCAGGGGC 1911

Right column:

Qy	2041	CGAGGAGGGGCGGGAAGTCAGCCAGGTTGCCAGAGAACACCGAGCCCGCCTCCCGCC 2100
Db	1912	CGAGGAGGGGCGGGAAGTCAGCCAGGTTGCCAGAGAACACCGAGCCCGCCTCCCGCC 1971
Qy	2101	CCCCAGGGCTTCTGCAGCCATCAGCCATTCACCATCCACCATCCGGATCTCTCCTGAA 2160
Db	1972	CCCCAGGGCTTCTGCAGCCATCAGCCATTCACCATCCACCATCCGGATCTCTCCTGAA 2031
Qy	2161	CTCCCACAGCGTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2220
Db	2032	CTCCCACGACGTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAAGCCAAGC 2091
Qy	2221	AAAATGCACACCCTTTTCTGTGGCAAATCGTCTGTCTGTACATGTGTACTAGTGTACATTAGAAA 2280
Db	2092	AAAATGCACACCCTTTTCTGTGGCAAATCGTCTGTCTGTACATGTGTACTAGTGTACATTAGAAA 2151
Qy	2281	GGGAAGATGTTAAGATATGTGGCCTGTGGGTGGGGTACCAGGGGTGCCTGCCAGCGGTAATATAT 2340
Db	2152	GGGAAGATGTTAAGATATGTGGCCTGTGGGTGGGGTACACAGGGGTGCCTGCCAGCGGTAATATAT 2211
Qy	2341	TTTAGAAATAATATATCAAATAACTCAACTAAACTCCAATTTTAATCAATTATTATTAATTTT 2400
Db	2212	TTTAGAAATAATATATCAAATAACTCAACTAAACTCCAATTTTAATCAATTATTATTAATTTT 2271
Qy	2401	TTTTTCTTTTAAAGAAAGAGGGCTTTTCTGTAGACTTTAAAGAAATAAAGTCTTTGGGAG 2460
Db	2272	TTTTTCTTTTAAAGAGAAGAGGGCTTTTCTGTAGACTTTAAAGAATAAAGTCTTTGGGAG 2331
Qy	2461	GTCTCACGGTGTAGAGAGGAGCTTTGAGGCGCCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
Db	2332	GTCTCACGGTGTAGAGAGGAGCTTTGAGGCGCCACCCGCACAAAAT TCACCCAGAGGGAAAT 2391
Qy	2521	CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACGAGAGGGGATA 2580
Db	2392	CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACGAGAGGGGATA 2451
Qy	2581	CCGTCTCCTTGAAGAGGAGAAACTCTGTCTGTCTCTAGCTCAGTCATACACCCATT 2640
Db	2452	CCGTCTCCTTGAAGAGGAGAAACTCTGTCTGTCTCTAGCTCAGTCATACACCCATT 2511
Qy	2641	TCTCTTTGCTTCACAGTTTTAAACTGTTTTTGCATATCTGCTATATAATTCTCTGTCT 2700
Db	2512	TCTCTTTGCTTCACAGTTTTAAACTGGTTTTTGCATATCTGCTATATAATTCTCTGTCT 2571
Qy	2701	CTCTCTGTTTTATCTCTCCCCTCCCCTCCCCTCCATTCTTTTGAA 2760
Db	2572	CTCTCTGTTTATCTCTCCCCTCCCCTCCCCTCCATTCTTTTGAA 2631
Qy	2761	TTTTCCCTCATCCCTCCATCTCAATCCGTATCCTCACGGCACCCCCCCCAGGCAAAGCA 2820
Db	2632	TTTTCCCTCATCCCTCCATCTCAATCCGTATCCTCACGGCACCCCCCCCAGGCAAAGCA 2691
Qy	2821	GTGCTCTGAGTATCACATCACACAGAGTCAATGTCAATTGTACTTGTCCTAGGCGTTTGGAAG 2880
Db	2692	GTGCTCTGAGTATCACATCACACAGAGTCAATGTCAATTGTACTTGTCCTAGGCGTTTGGAAG 2751
Qy	2881	CTTACACTTGGTTACTCCAACAACCACCAATCAACAAAGAAAAAATTCCAACAATG 2940
Db	2752	CTTACACTTGGTTACTCCAACAACCACCAATCAACAAAGAAAAAATTCCACAATG 2811
Qy	2941	AGGAAAACAGGAACCCACCAAAACAACAAGAACAACAATCAACAAAGAAAAATTCCAACATG 3000
Db	2812	AGGAAAACAGGAACCCACCAAAACAACAAGAACAACAATCAACAAAGAAAAAATTCCACAATG 2871
Qy	3001	AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
Db	2872	AAAGAATGTATTTTGTCTTTTGCATTTTGGTGTATAAGCCATCATATTCAGCAAAATG 2931
Qy	3061	ATTCCTTTCTTTAAAAAAAAAAAAAATTACCAAGGTTGTTGGGC 3120
Db	2932	ATTCCTTTCTTTAAAAAAAAAAAAAAATTACCAAGGTTGTTGGGC 2991
Qy	3121	CAGGGCGTTAAATTCACCAGATTTTTTTAACGAGAAAAACACACAGAGAGCTACCTCAG 3180

```
Db   61  AAGAGACCGATGATGAACAAGCTTTACATCGGGACAAGCTTGAGCCGTCGTCACCGCCGAC  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  GACTCCGGCAGCTCTTTGGGGACAGGAGCTGCCCCTGCCCTGGCCGGGGACAGTCTCGCTGATAG  180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GACTCCGGCAGCTCTTTGGGGACAGGAGCTGCCCCTGCCCTGGCCGGGGACAGTCTCGCTGATAG  180

Qy  181  TCCGGCTACGCCTTCGTGACTACCCGACGAACTGGCCATCCGCGCCATCGAGACGACC  240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCCGGCTACGCCTTCGTGACTACCCGACGAACTGGCCATCCGCGCCATCGAGACGACC  240

Qy  241  CTCTCGGTAAAGTGGAAGGTGACAGTCCAAGTCGTGCAACAGCTGTGTTCTTCCGCT  300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CTCTCGGTAAAGTGGAAGGTGACAGTCCAAGTCGTGCAACAGCTGTGTTCTTCCGCT  300

Qy  301  AAGCTAAGGAGCAGGAGAAATTCAGATTCAGATGTGGACAGTTGGAACAAGTCAACACAGAC  360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AAGCTAAGGAGCAGGAGAAATTCAGATTCAGATGTGGACAGTTGGAACAAGTCAACACAGAC  360

Qy  361  TTGGATGGACTTTTGGCTCAATATGGGACAGTTGGAGAATGTGGAACAAGAAGAAAATAGCCATG  420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  TTGGATGGACTTTTGGCTCAATATGGGACAGTTGGAGAATGTGGAACAAGAAGAAAATAGCCATG  420

Qy  421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAACAAGAAGAAGCAAAAATAGCCATG  480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAACAAGAAGAAGCAAAAATAGCCATG  480

Qy  481  GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT  540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GAGAAGCTAAGCGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGAT  540

Qy  541  GAAGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCAGCGTGGGAGACCACTCTTCCCG  600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GAAGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCAGCGTGGGACCACCACTCTTCCCGG  600

Qy  601  GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGCAGATTGATTCCCGCTGCGG  660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  GAGCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGCAGATTGATTCCCGCTGCGG  660

Qy  661  ATCCTGGTCCCCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG  720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  ATCCTGGTCCCCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAG  720

Qy  721  AACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT  780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  AACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAGAGAACTCTGGAGCT  780

Qy  781  GCAGAGAAGCCTGTCACCATCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATG  840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GCAGAGAAGCCTGTCACCATCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGCATG  840

Qy  841  ATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  ATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG  900

Qy  901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960

Qy  961  AAGAAAATTGAACATGAACATGAAACAGGAGGACCAAGATAACAATCTCATCTTTGAAAATGATATATGTGGCTTTGAGC  1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961  AAGAAAATTGAACATGAACATGAAACAGGAGGACCAAGATAACAATCTCATCTTTGAAAATGATATATGTGGCTTTGAGC  1020

Qy 1021  ATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCCTGTGCCAGTGCT  1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021  ATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCCTGTGCCAGTGCT  1080

Qy 1081  GGAGATAGAGATTATGAAGAAGCTGCGTGAGGAGAGCTCGCGTGAGAGCTCGCGTGTTAAC  1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081  GGAGATAGAGATTATGAAGAAGCTGCGTGAGGAGAGCTCGCGTGAGAGCTCGCGTGTTAAC  1140

Qy 1141  CAACCAAGCCAATCTGATCCCACAGGTTGAACCTCAGCGCCACTTGACCTCCATCTTTCAACAGGA  1200
```

```
Db  2992  CAGGGCGTTAAATTCACAGATTTTTTTAACGAGAGAAAACACACAGAGAAGAAGCTACCTCAG  3051
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181  GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052  GTGTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAG  3111

Qy  3241  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3171

Qy  3301  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTGTTCTGAATTGTCTTCCGCT  3360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTGTTCTGAATTGTCTTCCGCT  3231

Qy  3361  AGCCAAGAACCNATATGGCCTTCTTTTTGACAAACCTTGAAAATGTTTATTT  3412
             |||||||||||||||||||||||||||||||||||||||||||||
Db  3232  AGCCAAGAACCNATATGGCCTTCTTTTTGACAAACCTTGAAAATGTTTATTT  3283
```

RESULT 7
BD209927

LOCUS   BD209927   3283 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209927
VERSION   BD209927.1 GI:33019697
KEYWORDS   JP 2002512049-A/6.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3283)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE   Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same
  JOURNAL   Patent: JP 2002512049-A 6 23-APR-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT   OS   Homo sapiens (human)
            PN   JP 2002512049-A/6
            PD   23-APR-2002
            PF   16-MAR-1999 JP 2000545030
            PR   17-APR-1998 US 09/061709
            PI   YAO TSENG CHEN, ALI GURE, SOLAM TSANG, ELISABETH STOCKERT, ELKE
            JAGER,
            PI   ALEXANDER KNUTH, LLOYD J OLD
            PC   C12N15/09,A61K35/12,A61K39/00,A61K39/39,A61P35/00,C07K16/32,
            PC   C12N1/15,
            PC   C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
            574,
            PC   C12N15/00,C12N5/00
            CC   n=not determined(3243)
FH   Key   Location/Qualifiers
FT   source   1..3283
FT   /organism='Homo sapiens (human)'.

FEATURES   Location/Qualifiers
     source   1..3283
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxcn:9606"

ORIGIN

Query Match   92.1%; Score 3143; DB 6; Length 3283;
Best Local Similarity   96.2%; Pred. No. 0;
Matches 3283; Conservative   0; Mismatches   0; Indels   129; Gaps   1;

```
Qy     1  GGCAGGCGGAGGAGGCGGAGGAGCGCCGGGTACCGCGGGGAGGAGCCCGGGAGGAGCCGGGAGCCGCCGCCGCCGCCGAGCCGTCCTCCGGGG  60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GGCAGGCGGAGGAGGCGGAGGAGCGCCGGGTACCGCGGGGAGGAGCCCGGGAGGAGCCGGGAGCCGCCGCCGCCGCCGAGCCGTCCTCCGGGG  60

Qy    61  AAGAGACGGATGATGAACAAGCTTTACATCGGGACAAGCTTGAGCCGTCGTCACCCGCCGAC  120
```

```
Db  1141 ------------------------------------------------------------------  1140

Qy  1201 CTGTCCGTGTATCTCCACCAGCAGGGCCCGCGGAGCTCCCCCGCTCCCCCTACCAC  1260
Db  1141 ------------------------------------------------------------------  1140

Qy  1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCG  1320
              |||  ||||||||||||||||||||||||||||||||||||
Db  1141 ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCG  1191

Qy  1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAG  1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192 TTCCCCCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAG  1251

Qy  1381 GCTGTGGGCGCCATCACGGGAAGAAGGGGGCACACATCAAAACAGCTGCGAGATTGGCC  1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252 GCTGTGGGCGCCATCACGGGAAGAAGGGGGCACACATCAAAACAGCTGCGAGATTGGCC  1311

Qy  1441 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACTCAGCGAAAGGATGGTCATC  1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312 GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACTCAGCGAAAGGATGGTCATC  1371

Qy  1501 ATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA  1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372 ATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA  1431

Qy  1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAAGGCCATATCAGAGTGCCC  1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432 GAGGAAAACTTCTTTAACCCCAAAGAAGAGTGAAGCTGAAGGCCATATCAGAGTGCCC  1491

Qy  1621 TCTTCCACAGCTGCCGCGGTGATTGGCAAGGTGCAAGACCGTGACGAACTGCAGAAC  1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492 TCTTCCACAGCTGCCGCGGTGATTGGCAAGGTGCAAGACCGTGACGAACTGCAGAAC  1551

Qy  1681 TTAACCAGTTGCAGAATTATCGGGCACTTCTTTGCTAGCCAGAGATGAGGAAGTG  1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552 TTAACCAGTTGCAGAATTATCGGGCACTTCTTTGCTAGCCAGAGATGAGGAAGTG  1611

Qy  1741 ATCGTCAGAATTATCGGGCCACTGTCGTGCCTCGTGACCAGCGCAAGATCAGGGAA  1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1612 ATCGTCAGAATTATCGGGCCACTGTCGTGCCTCGTGACCAGCGCAAGATCAGGGAA  1671

Qy  1801 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTTCACAGCGC  1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1672 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTTCACAGCGC  1731

Qy  1861 AGCAAGTGAGGCTCCCACAGGCCACACAGGCGCGGCGGCCTCCGAGGCCCTGAGAACCCACAC  1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1732 AGCAAGTGAGGCTCCCACAGGCCACACAGGCGCGGCGGCCTCCGAGGCCCTGAGAACCCACAC  1791

Qy  1921 CTGACAGAATGAGACCAAAACGCAGCCAGCCGCAGATCGGAGGCAAACCAAAGACCATCTGAG  1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1792 CTGACAGAATGAGACCAAAACGCAGCCAGCCGCAGATCGGAGGCAAACCAAAGACCATCTGAG  1851

Qy  1981 GAATGAGAAGTTCTGCGGAGGGCGGGGAAGTCAGCCAGGTCAGCCAGCCTCCGCCTCCCGCC  2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1852 GAATGAGAAGTTCTGCGGAGGGCGGGGAAGTCAGCCAGGTCAGCCAGCCTCCGCCTCCCGCC  1911

Qy  2041 CGAGGAGGGGCGGGGAAGTCAGCCAGTTTGCCCAGGTTGCCAGCCCAGAACCCCGCCTCCCGCC  2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1912 CGAGGAGGGGCGGGGAAGTCAGCCAGTTTGCCCAGGTTGCCAGCCCAGAACCCCGCCTCCCGCC  1971

Qy  2101 CCCAGGGCTTCTGCGGGGGAAGGTCAGCCTTCAGCCATCCACTTCACCATCTCCTCCTGAA  2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1972 CCCAGGGCTTCTGCGGGGGAAGGTCAGCCTTCAGCCATCCACTTCACCATCTCCTCCTGAA  2031

Qy  2161 CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2032 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2091

Qy  2221 AAAATGCACACCCTTTTCTGTGTGCAAATCGTCTCTGTACATGTGTGTACATATTTAGAAA  2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2092 AAAATGCACACCCTTTTTCTGTGTGCAAATCGTCTCTGTACATGTGTGTACATATTTAGAAA  2151
```

```
Qy  2281 GGGAAGAGATGTTAAGAGATGTGGCTGTGGGTTACACAGGGTGCCTGCGAGCGGTAATATAT  2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2152 GGGAAGAGATGTTAAGAGATGTGGCTGTGGGTTACACAGGGTGCCTGCCTGCGAGCGGTAATATAT  2211

Qy  2341 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2400
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2212 TTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2271

Qy  2401 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2272 TTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2331

Qy  2461 GTCTCACGGTGTAGAGGAGGGAGCTTTGAGGCACCCGCACAAAAATTCACCCAGAGGGAAAT  2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2332 GTCTCACGGTGTAGAGGAGGGAGCTTTGAGGCCACCCGACAAAAATTCACCCAGAGGGAAAT  2391

Qy  2521 CTTCGTCGGAAGGACACTCACCGCAGTTCTGAATCACCTGTGATGTCAAGAGAAGAGGATA  2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2392 CTTCGTCCGAAGGACACTCACGGCAGTTCTGAATCACCTGTGATGTCAACAGAGAAGAGGATA  2451

Qy  2581 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATACACCCATT  2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452 CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATACACCCATT  2511

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTTGCTATATAATTCTCTGTCT  2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512 TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTTGCTATATAATTCTCTGTCT  2571

Qy  2701 CTCTCTGTTATCTCTCCCCTCTCCCCCTCTCCCCCCCCCCCCCCCCCAGGCAAAAGCA  2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2572 CTCTCTGTTATCTCTCCCCTCTCCCCCTCTCCCCCCCCCCCCCCCCCAGGCAAAAGCA  2631

Qy  2761 TTTCCTCATCCTCCATCTCAATCCGTATCCACACACGAACACAAAGCGAAACACCAGCCTCAA  2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632 TTTCCTCATCCTCCATCTCAATCCGTATCCACACACGAACACAAAGCGAAACACCAGCCTCAA  2691

Qy  2821 GTGCTCTGAGTATCACATCACAAGCAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAG  2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692 GTGCTCTGAGTATCACATCACAAGCAAGAGTCAATGTACTTGTCCTAGCGTTTTGGAAG  2751

Qy  2881 CTTACACTTGGTTACTCAAAGAAAGAACAAGAGTCAATATATTCAGCAAAATG  2940
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752 CTTACACTTGGTTACTCAAAGAAAGAACAAGAGTCAATATATTCAGCAAAATG  2811

Qy  2941 AGGAAAACAGGAACCACCAAACAACCAATCAACAAAGAAAAATTCCACAATG  3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812 AGGAAAACAGGAACCACCAAACAACCAACCAATCAACAAAGAAAAATTCCACAATG  2871

Qy  3001 AAAGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAAATG  3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2872 AAAGAATGTATTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAAATG  2931

Qy  3061 ATTCCTTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAAATTTACCAAGGTTGTTGGCC  3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932 ATTCCTTTCTTTAAAAAAAAAAAATGTGGAGGAAAGTAGAAAATTTACCAAGGTTGTTGGCC  2991

Qy  3121 CAGGGGGTTAAATTCACAGATTTTTTTTAACGAAGAAAAACACACAGAGATTTGTAAAGCTGATG  3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2992 CAGGGGGTTAAATTCACAGATTTTTTTTAACGAGAGAAAAACACACAGATTTTGTAAAGCTGATAG  3051

Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTGTCTCTTGTGTCTTTTAATAAAAATGAGTGGAAAAAATAAGATATCAACT  3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052 GTGTTTTTACCTCAGCACCTTGCTCTGTCTCTTGTGTCTTTTAATAAAAATGAGTGGAAAAAATAAGATATCAACT  3111

Qy  3241 TTGGAGCATTTTTTATTTTTTATTTTTTTAATAAAAAAATGAGTGCACAACAGTGTGCAACAGTGTTCTTCCGGT  3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112 TTGGAGCATTTTTTTATTTTTTATTTTTTTAATAAAAAAATAAGATGCACAACAGTGTGCAACAGTGTTCTTCCGGT  3171

Qy  3301 GCCAGCCTGAGAAGTGACAGTGCAAGTGCACAACAGTGTTCTTGAATTGTCTTCCGCT  3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172 GCCAGCCTGAGAAGTGACAGTGCAACAGTGTTCTTGAATTGTCTTCCGCT  3231
```

```
Qy   3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          ||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3232 AGCCAAGAACNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

RESULT 8
AF057352

LOCUS   AF057352   3667 bp   mRNA   linear   PRI 23-MAY-1999
DEFINITION Homo sapiens hepatocellular carcinoma autoantigen (p62) mRNA,
            complete cds.
ACCESSION   AF057352
VERSION   AF057352.1 GI:4883680
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3667)
  AUTHORS   Zhang,J.Y., Chan,E.K., Peng,X.X. and Tan,E.M.
  TITLE   A novel cytoplasmic protein with RNA-binding motifs is an
            autoantigen in human hepatocellular carcinoma
  JOURNAL   J. Exp. Med. 189 (7), 1101-1110 (1999)
  MEDLINE   99207072
   PUBMED   10190901
REFERENCE   2 (bases 1 to 3667)
  AUTHORS   Zhang,J.Y., Chan,E.K.L., Peng,X.X. and Tan,E.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-APR-1998) Molecular & Experimental Medicine, The
            Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla,
            CA 92037, USA
FEATURES   Location/Qualifiers
     source   1..3667
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene   1..3667
                     /gene="p62"
     CDS   436..2106
                     /gene="p62"
                     /note="RNA-binding protein"
                     /codon_start=1
                     /product="hepatocellular carcinoma autoantigen"
                     /protein_id="AAD31596.1"
                     /db_xref="GI:4883681"
                     /translation="MMNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVD
                     YPDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLL
                     AQYGTVENFEQVNTDETAVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEV
                     SSPSPPQRAQRGDHSSREGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKGLTIKN
                     ITKQTQSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIP
                     LKILAHNGLVGRLIGKEGRNLKKIEHETGTKITISSLQDLSIYNPRTITVKGTVEAC
                     ASAEIEIMKKLREAFENDMLAVNTHSGYFSSLYPHHQFGPFPHHHSYPEQEIVNLFIP
                     TQAVGAIIGKKGAHIKQARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIF
                     GKLKEENFPNPKEEVKLEAHIRVPSSTAGRVIGKGKTVNELQNLTSAEVIVPRDQTP
                     DENEEVIVRIIGHFASQTAQRKIREIVQQVKQEQKYPQGVASQRSK"
```

ORIGIN
Query Match       91.3%;  Score 3114.8;  DB 9;  Length 3667;
Best Local Similarity   96.1%;  Pred. No. 0;
Matches 3277; Conservative   0; Mismatches   3; Indels 131; Gaps   3;

```
Qy      3 CAGCGGAGGAGGCGAGGAGCCCGGTACCGGCGGAGCCGGGAGCCTCGGGGAA 62
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CAACGGAGGAGGCGAGGAGCCCGGTACCGGCGGAGCCGGGAGCCTCGGGGAA 428

Qy     63 GAGACGGATGATGAACAAGCTTACATCGGAAGCTTACATCGGAAGCCGCACGA 122
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GAGACGGATGATGAACAAGCTTTACATCGGAACAAGCTTACATCGGAAGCCGCACGA 488

Qy    123 CCTCCGGCCAGCTTCTTTGGGACAGGAAGCTGCCCCTGGGGGGGACAGGTCCTGTCAAGTC 182
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CCTCCGGCCAGCTTCTTTGGGACAGGAAGCTGCCCCTGGCCCCGGCGGGGACAGGTCCTGTCAAGTC 548

Qy    183 CGGCTACGGCGCCTTCGTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGAGACCCTT 242
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    549 CGGCTACGGCCTTCGTGGACTACCCCGACCAGAAACTCATGGGACCATCCGCGCCATCGAGACCCT 608

Qy    243 CTCGGGTAAAGTGAATTGCATGGAAAATCATGGAAGTTGATTACTCAGTTCTCTAAAAA 302
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 CTCGGGTAAAGTGAATTGCATGGAAAATCATGGAAGTTGATTACTCAGTTCTCTAAAAA 668

Qy    303 GCTAAGGAGGCAGAAATTCAGATTGAAACTCCCTCCTGCCAGTGGAGGTGTT 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GCTAAGGAGGCAGGAAATTCAGATTGAAACATCCCTCCTCACCTGCCAGTGGGAGGTGTT 728

Qy    363 GGATGGACTTTTGGCTCAATATGGACAGTGGAACAGTGGGAGAATGTGGAACAGTGGAACACAGAC 422
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GGATGGACTTTTGGCTCAATATGGACAGTGGGAGAATGTGGAACAGTGGAACACACAGAC 788

Qy    423 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAGAGCAAAAATAGCCATGGA 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGA 848

Qy    483 GAGCTAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA 542
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA 908

Qy    543 AGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGA 602
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 AGAGGTGAGCTCCCCTTGCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCCGGGA 968

Qy    603 GCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGAATTCCCGCTGCGGAT 662
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GCAAGGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGAATTCCCGCTGCGGAT 1028

Qy    663 CCTTGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTTGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 1088

Qy    723 CATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAGAAACTCTGAGCTGC 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAGAAAGAACTCTGAGCTGC 1148

Qy    783 AGAGAAGCTGTCACCATCCATGCACCCCAGAGGGGACTTCTGAAGCATGCGGCATGAT 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCTGTCACCATCCATGCACCCCAGAGGGGACTTCTGAAGCATGCGGCATGAT 1208

Qy    843 TCTTGAAAATCATGCAGAAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCGAA 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAAATCATGCAGAAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCGAA 1268

Qy    903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAA 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAA 1328

Qy    963 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCAT 1022
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTGAGCAT 1388

Qy   1023 ATACAACCCGGAAAGAACACCACTCGTGAAGGCCACAGTGCCAGTGCTGA 1082
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGGAAAGAACAACCATCACTGTGAAGGCCACAGTGCCAGTGCTGA 1448

Qy   1083 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATAGTGCTGGCTGTTAACCA 1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATAGTGCTGGTTGTTAAC- 1506

Qy   1143 ACAAGCCAATCGATCCCAGAGTTGAACCTGCCACTTGGACATCTTTCAACAGGACT 1202
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------- 1506

Qy   1203 GTCCGTGCTATCTCCACCAGACAGGGCCCCGGAGCTCCCCCGCCCGCTACCACCC 1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------------------------------------------------- 1506

Qy   1263 CTTCACTACCCCACTCCGGATACTTCTCCAGCTGTGACCCCATCACCAGTTTGGCCCGTT 1322
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

Left column (alignment blocks):

```
Db   1507 ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTT 1559

Qy   1323 CCGGCATCATCACTCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCAGGC 1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CCCGCATCATCACTCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGC 1619

Qy   1383 TGTGGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG 1442
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1620 TGTGGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGG 1679

Qy   1443 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGACGTCAGCGAAAGATGGTCATCAT 1502
          |||||||||||||||||||||||||||||||||||||
Db   1680 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT 1739

Qy   1503 CACCGGGCACCCGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTTGGGAAACTGAAAGA 1562
          ||||||||||||||||||||||||||||||||||||||
Db   1740 CACCGGGCACCCGAAGCCCAGTTCAAGGCCCAGGGACGCGGATCTTTGGGAAACTGAAAGA 1799

Qy   1563 GGAAAACTTCTTTAACCCCAAAGAGAGAAGTGAAGCTGGAAGCTGAAGTCAGAGTGCCCTC 1622
          ||||||||||||||||||||||||||||||||||||
Db   1800 GGAAAACTTCTTTAACCCCAAAGAGAGAAGTGAAGCTGGAAGCGCATAGCAGAGTGCCCTC 1859

Qy   1623 TTCCACAGTCTGCCGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTT 1682
          ||||||||||||||||||||||||||||||||||
Db   1860 TTCCACAGTCTGGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTT 1919

Qy   1683 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT 1742
          |||||||||||||||||||||||||||||||||
Db   1920 AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT 1979

Qy   1743 CGTCAGAATATCGGGCACTCTTCTTTGCTCGTAGCCCAGATGCACAGGCAAGATCAGGGAAAT 1802
          ||||||||||||||||||||||||||||||||||
Db   1980 CGTCAGAATTATCGGGCACTCTTCTTTGCTCGTAGCCCAGATGCACAGCCAAGATCAGGGAAAT 2039

Qy   1803 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG 1862
          ||||||||||||||||||||||||||||||||||
Db   2040 TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG 2099

Qy   1863 CAAGTGAGGTCCACAGGACACCAGCAGCAAAACAACGATGAAATGTAGCCCTTCCAACACCT 1922
          ||||||||||||||||||||||||||||||||
Db   2100 CAAGTGAGGTCCACAGGACACCAGCAGCAAAACAACGATGAAATGTAGCCCTTCCAACACCT 2159

Qy   1923 GACAGAATGAGACCAAAACGCAGCCAGATCGGAGCAAACCAAGACCATCTGAGGA 1982
          |||||||||||||||||||||||||||||||
Db   2160 GACAGAATGAGACCAAAACGCAGCCAGATCGGAGCAAACCAAACAAAGACCATCTGAGGA 2219

Qy   1983 ATGAGAAGTCTGCGGAGGGCGGCCAGGGGCGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCG 2042
          |||||||||||||||||||||||||||||||||
Db   2220 ATGAGAAGTCTGCGGAGGGCGGCCAGGGGCGCCAGGGACTCTGCCGAGGCCCTGAGAACCCCAGGGGCCG 2279

Qy   2043 AGGAGGGGCGGGGAAGGTCAGCGAGTTGCCAGGGTTGCCAGAACCACCGGCCTTCCCGCCTTCCCCCC 2102
          |||||||||||||||||||||||||||||||
Db   2280 AGGAGGGGCGGGGAAGGTCAGCGAGTTGCCAGGGTTGCCAGAACCACCGGCCTTCCCGCCTTCCCCCC 2339

Qy   2103 CCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2162
          ||||||||||||||||||||||||||||||||
Db   2340 CCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2399

Qy   2163 CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2222
          |||||||||||||||||||||||||||||||
Db   2400 CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2459

Qy   2223 AATGCACACCCTTTTCTGTGGTACCATTGGCCTGGCCGTGCCTGCTGCAGCGGTAATATATTT 2282
          ||||||||||||||||||||||||||||||||
Db   2460 AATGCACACCCTTTTCTGTGGTACCATTGGCCTGGCCGTGCCTGCTGCAGCGGTAATATATTT 2519

Qy   2283 GAAGATGTTAAGATGTGTGCCCTGGGTTACACAGGGGTACACAGGGGTAATATATTT 2342
          ||||||||||||||||||||||||||||||
Db   2520 GAAGATGTTAAGATGTGTGCCCTGGGTTACACAGGGGTACACAGGGGTAATATATTT 2579

Qy   2343 TAGAAATAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTAATTAATTTTTT 2402
          |||||||||||||||||||||||||||||
Db   2580 TAGAAATAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTAATTAATTTTTT 2639
```

Right column (alignment blocks):

```
Qy   2403 TTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2462
          |||||||||||||||||||||||||||||||||||||||||||
Db   2640 TTTCTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2699

Qy   2463 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCAGAGGGAAATCT 2522
          |||||||||||||||||||||||||||||||||||
Db   2700 CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCAGAGGGAAATCT 2759

Qy   2523 CGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTGTATGTCAACAGAAGGGATACC 2582
          |||||||||||||||||||||||||||||||||
Db   2760 CGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTGTATGTCAACAGAAGGGATACC 2819

Qy   2583 GTCTCCTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCGTAGCTCATCACCCCATTTC 2642
          |||||||||||||||||||||||||||||||
Db   2820 GTCTCCTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCGTAGCTCATCACACCCATTTC 2879

Qy   2643 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCGTCTCT 2702
          |||||||||||||||||||||||||||||||||
Db   2880 TCTTTGCTTCACAGGTTTCACAGGTTTTTTGCATACTGCTATATAATTCTCGTCTCT 2939

Qy   2703 CTCGTTTATCTCTCCCCCTCCTCCCCCTTCTTCTCCATCTCCATTCTTTTGAATT 2762
          |||||||||||||||||||||||||||||||
Db   2940 CTCGTTTATCTCTCCCCCTCCTCCCCCTTCTTCTCCATCTCCATTCTTTTGAATT 2999

Qy   2763 TCCTCATCCCTCCATCTCAATCCGTATCTACGGCACCCCCCCCCCCCAGGCAAAGCAGT 2822
          ||||||||||||||||||||||||||||||
Db   3000 TCCTCATCCCTCCATCTCAATCCGTATCTACGCA-CCCCCCCCCCCCAGGCAAAGCAGT 3058

Qy   2823 GGTCTGAGTATCACATCACACAAAAGGCGAAACACACAAAACCAGCCTCAACT 2882
          |||||||||||||||||||||||||||||
Db   3059 GGTCTGAGTATCACATCACACAAAAGGCGAAACACACAAAACCAGCCTCAACT 3118

Qy   2883 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTCGTCCTAGCGTTTTGGAAGAG 2942
          |||||||||||||||||||||||||||||||
Db   3119 TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTCGTCCTAGCGTTTTGGAAGAG 3178

Qy   2943 GAAAACAGGAACCACCAAACCAATCAACCAACAAACAAGAAAAATTCCACAATGAA 3002
          |||||||||||||||||||||||||||||||
Db   3179 GAAAACAGGAACCACCAAACCAATCAACCAACAAACAAGAAAAATTCCACAATGAA 3238

Qy   3003 AGAATGTATTTTGTCTTTTGTCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3062
          |||||||||||||||||||||||||||||||
Db   3239 AGAATGTATTTTGTCTTTTGTCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3298

Qy   3063 TCCTTTCTT-AAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGCCCC 3121
          ||||||||| ||||||||||||||||
Db   3299 TCCTTTCTTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGCCCC 3358

Qy   3122 AGGGCGTTAAATTCACAGATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG 3181
          |||||||||||||||||||||||||||||||
Db   3359 AGGGCGTTAAATTCACAGATTTTTTAACGAAAATGAGTTGGAAAAAAACACAGAAGAGTACCTCAGG 3418

Qy   3182 TGTTTTTACCTCAGCACTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3241
          |||||||||||||||||||||||||||||||
Db   3419 TGTTTTTACCTCAGCACTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGCTGATAGT 3478

Qy   3242 TGGAGCATTTTTTATTTTTAATAAAAATGAGTGTCAACAGTTGCCAAGTGTGCAACAGTCCAAGT 3301
          |||||||||||||||||||||||||||||||
Db   3479 TGGAGCATTTTTTATTTTTAATAAAAATGAGTGTCAACAGTTGCCAAGTGTGCAACAGTCCAAGT 3538

Qy   3302 CCAGCCTGGAGAAGGTGACAGTCCAAGTCGTCAACAGCTGTTCTTCCGCTA 3361
          |||||||||||||||||||||||||||||
Db   3539 CCAGCCTGGAGAAGGTGACAGTCCAAGTCGTCAACAGCTGTTCTTCCGCTA 3598

Qy   3362 GCCAAGAACNATATGCCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3412
          |||||||||||||||||||||||||||
Db   3599 GCCAAGAACCTATATGCCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3649
```

RESULT 9
AC020629
LOCUS   AC020629   142971 bp   DNA   linear   PRI 07-MAR-2002

DEFINITION   Homo sapiens 12q BAC RP11-76E16 (Roswell Park Cancer Institute
              Human BAC Library) complete sequence.
ACCESSION   AC020629
VERSION   AC020629.6 GI:7656675
KEYWORDS   HTG.
SOURCE   Homo sapiens (human).
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 142971)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
              Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
              Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
              David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
              Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
              Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
              Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
              Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
              Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
              Lau,S., Leal,B., Lee,E., Li,J., Lichtarge,O., Liu,J., Liu,W.,
              Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
              Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A.,
              Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
              Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
              Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
              Samuel,S., Say,J., Scherer,S., Shah,P., Shen,H., Shim,C., Simon,M.,
              Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
              Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
              Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
              Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
              Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 142971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 142971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-APR-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 142971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (28-APR-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5 (bases 1 to 142971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Apr 27, 2000 this sequence version replaced gi:7025656.
              INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
              gc-help@bcm.tmc.edu

              CLONE LENGTH: This sequence does not necessarily represent the
              entire insert of this clone. Overlapping regions of clones are only
              sequenced and submitted once, so the sequence for the remainder of
              the insert may be found in the record for the adjacent clones.
              Overlapping clones are noted at the beginning and end of the
              Features listing.

              ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550) searches
              of a local database that includes entries from dbSTS, GDB, and
              local mapping efforts.
              Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT--------

----------- Summary Statistics ---------------

Contig length: 142971
Phrap values in estimate: 140680
Average error rate (BCM-Phrap estimate):   0.000118703
Fraction of Phrap values less than 40 :   0.0433608
Number of consensus changing edits: 31
Number of N's in consensus :   0

------------- Consensus changing edits --------
Position   Original+Context   Edited+Context
11171   gtcttctgga(n)caaaaaaaa   gtcttctgga(a)caaaaaaaa
12018   tactcatctg(n)ctatatctat   tatccatctg(t)ctatatctat
16604   tccactctaa(n)agttccattt   tccactctaa(g)agttccattt
18385   tagtaaccac(n)tacattttta   tagtaaccac(t)tacattttta
21879   aaaaaaaaaa(a)aattaaaaca   aaaaaaaaaa(a)aattaaaaca
21880   aaaaaaaaan(n)nttaaaaacag   aaaaaaaaaa(a)attaaaaacag
21881   aaaaaaaann(n)ttaaaaacagg   aaaaaaaaan(a)ttaaaaacagg
21906   cctaggtccc(n)cttaaatgag   cctaggtccc(a)cttaaatgag
22169   accagcctgg(n)caacatggcg   accagcctgg(c)caacatggcg
38915   ttacatctag(n)tcttttgcaa   ttacatctag(t)tcttttgcaa
55675   acttgtctc(n)aaaaaaaaaa   acttgtctc(a)aaaaaaaaaa
57658   gaccagcctg(n)ccaacatggt   gaccagcctg(a)ccaacatggt
74371   gtgtcggccc(n)ccgccgcgc   gtgtcggccc(c)ccgccgcgc
88021   taaatgaggg(n)agaccagtgg   taaatgaggg(c)agaccagtgg
88248   gcagtgagcc(n)agatcacacc   gcagtgagcc(a)agatcacacc
94121   ttgttcccga(n)gtctcctcgc   ttgttcccga(t)gtctcctgc
95228   gtccatggt(n)cnnttgttc   gtccatggt(t)ctcattgttc
95230   ccatgtgtc(n)nnttgttcaa   ccatgtgtc(t)cattgttcaa
95231   cagtgtncn(n)nttgttcaact   cagtgtgtcc(c)attgttcaa
95232   atgtgtncnn(n)ttgttcaact   atgtgttctc(a)ttgttcaact
104931   ttttttttag(n)acagacagg   ttttttttag(t)acagacagg
118168   gcccaggct(n)ccaaggtgt   gcccaggtcc(c)ccaaggtt
120974   catctcaaaa(n)ananangaac   catctcaaaa(a)aaaaagaac
120976   tctcaaaaaa(n)anangaactc   tctcaaaaaa(a)aaaagaactc
120980   tcaaaaaana(n)angaactctc   tcaaaaaaaa(a)angaactctc
120980   aaaaanana(n)agaactcttcc   aaaaaaaaaa(a)gaactcttcc
139505   ccagcctgg(n)ggggggcacc   ccagcctgg(t)ggtgggcacc
139620   gccatgggg(n)nggggcacctgt   gccatggg(t)ggcacctgg
139698   attaccatta(n)aatatggtat   attaccatta(g)aatatggtat
140104   accgtgccca(n)cagaatagaa   accgtgccca(g)cagaatagaa
140149   tggaacaatt(n)gataccaca   tggaacaatt(g)gataccaca

--------- Distribution of Quality < 40 Bases ---------

                                            * *

                                            * *

             1000|
              900|

```
Qy      241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA   300
             ||||||||| |||||||||||||||||||||||||||||||||||||| | |||||||| |
Db   133193  CTCTCGGGTCAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATATTCAGTCTCTATA   133252

Qy      301  AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGAGAGGTG   360
             |||||||||||||||||||| |||||||||||| ||||||||| ||||||||||||||||
Db   133253  AAGCTAAGGAGCAGGAGAACATTCGATTCGAAATATCCCTCCTCACCTGCAGTGGAGAGGTG  133312

Qy      361  TTGGATGGACTTTGGTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGAC     420
             |||||||||||||||| |||||||| | |||||||||||||| |||||||||||||||
Db   133313  TTGGATGGACTTTGGCTCAATATGGGCTGAGAATGTGGAGCAAGTCAACACAGAC     133372

Qy      421  ACAGAAACCGCCGTTGTCAAGTCACATATGCAACAGAGAAGAAGTAAAAATAGCCATG    480
             ||||| | ||| |||||||||||||||||||| |||||||||||||||||||||||||
Db   133373  ACAGAGACGCTGTTGTCAAGTCACATATGCACATATGCAACAAAGAGAAGTAAAAATAGCCATG  133432

Qy      481  GAGAAGCTAAGGGGCATCAGTTGAGAACTACTCCTTCAGATTTCCTACATCCCGGAT    540
             ||||||||||| ||||||||||||||||||| |||| |||||||||| ||| |||||
Db   133433  AAGAAGCTAAGCGGCATCAGTTTGAGAACCACTACTTCAGATTTCAAGATTTCCTACATCCCGGAT  133492

Qy      541  GAAGAGTGAGCTCCCCTTCGCCCCTCCCCCTCAGCGAGCCCAGGTGGGGACCTCTTCCCGG   600
             |||||||||||||||| ||||||||| ||||||||||||||||||||| |||| || |||
Db   133493  GACGAGTGAGCTGCCCTTCGCCCCTCCCCCTCAGCGAGCCCAGGTGGGGACCACTCTTCCTGG  133552

Qy      601  GGCAAGGCCACGCCCCTGGGGGGCCAGAGCAGATTGATTTCCCGTGCGG    660
             |||||||||||||||||||||||||||||||||| |||||||||| |||
Db   133553  GGCAAGGCCAAGCCCCTGGGGGGCTCTTCTCAGGCCAGACAGATTGATTTCCCACTGCGT  133612

Qy      661  ATCCTGGTCCCCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG   720
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   133613  GTCCTGTTCCCCACCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG  133672

Qy      721  AACATCACTAAGCAGACACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT   780
             |||||||| |||||||||||||||| |||||||| ||||| |||||||||||||| ||||||
Db   133673  AACATCACTAAGCAGACCGGCGTCCCCGGGTAGCATCTTATACGACAAGAGAACTCCCAGAGCT  133732

Qy      781  GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGGACTTCTGAAGCATGCCGCCATG   840
                                ||||||||||||||||||||| ||||||||||||||||||||
Db   133733  GCAGAGAAGCCTGTCA----CCATGCCACCCCAGAGGGGGACTTCTGAAGCATGCCGCCATG  133788

Qy      841  ATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG   900
             ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   133789  ATTCTTTGAAATAATGCAGAAGAGGCAGATGAGGCCAAACTAGCCGAAGAGATTCCTCTG  133848

Qy      901  AAAATCTTGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG   960
             ||||||||||||||| ||||||||| |||||||| ||||||||||||||||||||||||
Db   133849  AAAATCTTGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  133908

Qy      961  AAGAAATTGAACATGAAAACAGGGACCAAGATAACAATCTTCGAGGATTTGAGCC   1020
             |||||| ||||||||| ||||||||||||||| ||||||||||||||||||||||
Db   133909  AAGAAAAATGAACATGAAAACAGGGACCAAGATAACAATCTTCGCAGGATTTGAGC  133968

Qy     1021  ATATACAACCGGAAAGAAGAACCATCACTGTGAAGGGCACAGTTGAGCCGTCCAGTGCT   1080
             |||||||||||||||||||| ||||||||||||||| |||||||||||||| ||||||||
Db   133969  ATATACAACCCCGGAAAGAAGAACCATCACTGTGAAGGGCACAGTGCTGTGCGCCAGTGCT  134028

Qy     1081  GAGATAGAGATTATGAAGAACGCTGCCTGAGCCTTTGAAAATGATATGCTCGCTGTTAAC   1140
             ||||||||||||||||||||| ||||| |||||||||||||||||| |||||||||||||
Db   134029  GAGATAGAGATTATGAAGAAGCGCTGCCTGAGCCTTTGAAAATGATACGCTGACTGTTAA   134087

Qy     1141  CAACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGCAGTTCTTTTCAACAGGA   1200
                                                      
Db   134088  ------------------------------------------------------  134087

Qy     1201  CTGTCCGTGCTATCTCCACCAGCAGCAGGCGCCCTGGCTGAAG   1260
                                                         
Db   134088  -------------------------------------------  134087

Qy     1261  CCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG   1320
                                      ||||||||||||||||| |||||||||||||||||
Db   134088  ------------TACCCACTTCGGATACTTCTCCAGCCTGTACCCCATCCGCCAGTTTGGCCCG  134139
```

```
         800  ──
         700  ──
# bases  600  ──
         500  ──
         400  ──                          *  *
         300  ──                *  *  *  *  *  *
         200  ──             *  *  *  *  *  *  *  *
         100  ──    *  *  *  *  *  *  *  *  *  *  *  *  *  *
           0  ──  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
                   5   10   15   20   25   30   35   40
                          Phrap Value Range
```

FEATURES   Version: 1.01 qxfo.
            Location/Qualifiers
source   1..142971
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-76F16"
repeat_region   147..236
            /rpt_family="A-rich"
repeat_region   258..378
            /rpt_family="AluJb"
repeat_region   complement(537..840)
            /rpt_family="AluSx"
repeat_region   complement(1722..2137)
            /rpt_family="Tigger2a"
repeat_region   complement(3161..3584)
            /rpt_family="MSTD"
STS   3228..3502
            /standard_name="G05956"
            /db_xref="dbSTS:11453"
STS   3347..3490
            /standard_name="cda0yb04"
            /db_xref="GDB:446196"
misc_feature   3585..3735
            /note="Region similar to Hs#S936143 aj27d02.s1 Homo
            sapiens cDNA: AA789222"
repeat_region   3885..3910
            /rpt_family="(TA)n"
repeat_region   complement(3917..4030)
            /rpt_family="AluJb"
repeat_region   complement(4544..4681)
            /rpt_family="MIR"
misc_feature   5145..5510
            /note="Region similar to Hs#S953417 oa06g05.s1 Homo
            sapiens cDNA: AA765439"
misc_feature   5553..5859
            /note="Region similar to Hs#S1195593 qt69f10.x1 Homo

Query Match   78.0%;   Score 2662; DB 9; Length 142971;
Best Local Similarity 90.2%;   Pred. No. 0;
Matches 3085; Conservative 0; Mismatches 156; Indels 180; Gaps 15;

```
Qy        1  GGCAGCGGAGGAGGCGAGGAGCGCCGGCCGCCGGGAGCCGCGGGGGCCTCGGGG    60
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   132954  GGCAGCGGAGGAGGCGAGGAGCGCCGGCCGCCGGGAGCCGCGGGGGCCTCGGGG  133012

Qy       61  AAGAGACGGATGATGAACAAGCTTACATCGGAGAACCTGAGCCCGCCGTCACCGGCGGAC   120
             |||||||||||||| ||||| ||||||||||||||||||||||||| |||||||||||||
Db   133013  AACAGACGGATGATGAACAAGCTTTTCATCGGGAGAACCTGAGCCCGTCACCGCGCCGGAA  133072

Qy      121  GACCTCCGGCAGCTCTTTGGGGACAGGAAGAGCTGCCCCGCGCGGGGGACAGGTCCTGCTGAAG   180
             ||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db   133073  GACCTCCGGCAGCTCTTTGGGGACAGGAAGAGCTGCCCCGCGCGGGGGACAGGTCCTTGCTCAAG  133132

Qy      181  TCCGGCTACGGCCTTCGTGTGGACTACCCCGACCACCCCGAAACTCGCCGCCATCCGGCGAGACC   240
             |||||||||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db   133133  TCCCGCTACGGCCTTCGTGTGGACTACCCCGACCACCCCCATCCGCCCCATCCGGACCATCGAGACC  133192
```

```
Qy     1321 TTCCCGCATCATCACTCTTATCCAGACGAGGAGATTGTGAATCCTTCATCCCAACCCAG 1380
            ||||||||| || ||||| || ||||||||||||||||||| ||| ||| |||||||||
Db   134140 TTCCCGCATCATCACTCTTATCCAGACGAGGAGATTGTCATCAATCTTCATCCCAACCCAG 134199

Qy     1381 GCTGTGGGCGCCCATCATCGGGGAAGAAGGGGCACACATTCAAACAGCTGGCGAGATTCGC 1440
            ||||||||| |||||||||||||| |||||||| ||||||||||||||| |||||||| ||
Db   134200 GGTGTGGGGCCCATCATCGGGGAAGAAGAAAGGGGCACACATCAAACAGCTGGCGAGATTCGTG 134259

Qy     1441 GGAGCCTTATCAAGATTGCCCCTTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500
            ||||||||||||||||||||||||||||||||||||||||||            ||||||
Db   134260 GGAGCCTCCATCAAGATCGCCCCTGCGGTCAG------------CGGAAGGTCATC 134302

Qy     1501 ATCACCGGGCACCCGGAAGCCAGTTCAGGGACGGATCTTTGGGAAACTGAAA 1560
            |||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   134303 ATCACCTGCCACCCCGGAATCCCAGGGACCCCAGGGACGGATCTTTGGGAAACTGAAA 134362

Qy     1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAAGCGTGAAGCTGGAAGCGCATATCGAAGTGCCC 1620
            |||||||||||||||||||||||||||||||| ||||||||||||||| |||||||||||
Db   134363 GAAGAAAACTTTTTTAACCCCAAAGAAGAAGCGTGGAAGCTGGAAGCGCTGAAACCCATATCAGAGTGCCC 134422

Qy     1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1680
            |||||||||||||| ||||||||||| |||||||||||| |||||| |||||||||||
Db   134423 TCTTCCACCGCTGGCCGGGATGATTGGCAAGGGGCAAGACCGTGAATGAACTGCAGAAT 134482

Qy     1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACCCAGATGAAAATGAGGAAGTG 1740
            || ||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   134483 TTAATCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACCCAGATGAAAATGAGGAAATG 134542

Qy     1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAAGTGCACAGCGCAAGATCAGGGAA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134543 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAAGCGCACAGCGCAAGATCAGGGAA 134602

Qy     1801 ATTGTACACAGGTGAAGCAGCAGGAGCAGAGAAATCCCTCAGGGAGTCGCCTCACAGCGC 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134603 ATTGTACAACAGGTGAAGCAGCAGGAGCAGAGAAATCCCTCAGGGAGTCGCCTCACAGCGC 134662

Qy     1861 AGCAAGTGAGGGCTCCCACAGGCGACCAAGCAAGACAAGCAAACACAAGTGAGCCCTCCAACAC 1920
            ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   134663 AGCAAGTGAGGATCTGCGGGGGAAGACCAAGCAAGACAAGCAAACACAAGTGAGGAAGTGTTATACAAG 134722

Qy     1921 CTGACAGAATGAGAGACCAAAGCCAGCCAGATCGGGAGCAAACCAAGACCATCTGAG 1980
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134723 CTGACAGAATGAGAGACCAAAGCCAGCCAGATCAAGAGCAAACCAAGACCATCTGAG 134782

Qy     1981 GAATGAGAAGTCTCGGCGGAGGCGGAAGGTTCAGCCAGGTTTGCCAGACACCCGAGAACCCCAGGGGC 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134783 GAATGAGAAGTCTCGGCGGAGGCGGAAGGTTCAGCCAGGTTTGCCAGACACCCGAGAACCCCAGGGGC 134842

Qy     2041 CGAGGAGGGGCGGCGGGGGGAAGGTAAACCAGGTTGCCAGAACCACCGGCCCCTCCCTCCTGTCC 2100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134843 CGAGGAGGGGCGGCGGGGGGAAGGTAAACCAGGTTGCCAGAACCACCGGCCCCTCCCTCCTGTC 134902

Qy     2101 CCCCGGGGCTTCTGCAGGCTTCAGCCATCCACTTGAACTAACATAGGTGAAC-GTGTTCAAAGCCAAG 2160
            ||||||||||||||||||||||| ||||||||||||||||||||||| |||||||||||||
Db   134903 CCCCATGACGCTATCCCTTTTAGTTGAACTAACATAGGTAAACATTGTTCAAAGCCAAG 134962

Qy     2161 CTCCCACGACGCTATCCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAA 2219
            ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   134963 CTCCCATGACGCTATCCCTTTTTCTGTGGCAAATCGTCTCTGTACATTGTTGTACATATTAGAA 135022

Qy     2220 CAAAATGCACACCCTTTTTT-TGTGGCAAATAGTCTTCACACGGTGTATACATATTAGGA 2279
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   135023 CAAAATGCACACGCCCTTTTTTGTGGCAAATAGTCTTCACACGGTGTATACATATTAGGA 135081

Qy     2280 AGGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCAGCGGTAATATA 2339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135082 GGGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCAGCGGTAATATA 135141

Qy     2340 TTTTAGAAATAATATCAAATAACTAACTCCAATTTTTAATCAATTATTAATTT 2399
            ||||||||| |||| |||| ||| ||||||||||||| |||||| |||  |
Db   135142 TTTTAGAAATATATCAAATAACAAT-----AACTAACTCCAATTTTTAATCAATTATTA---T 135193

Qy     2400 TTTTTTTCTTTTTAAAGAGAAGAAAGCAGGCTTTTCGAGACTTTTAAAGAATAAAAGTCTTTGGGA 2459
            |||||||||||||||||||||||||||||  |||||||||||||||| ||  |||||||||||
Db   135194 TTTTTTCTTTTTAAAGAGAGAAGAGCAGGCTTTTCGAGACTTTAAAGTCT---TTCTTTAGAA 135250

Qy     2460 GGTCTCACGTGTGAGAGAGGAGGCAGGCTTTGAGGCCACCCGACACAAAATTCACCCAGAGGAGAAA 2519
            ||||||||||||||||||||||||||||||||||||||| ||||| |||||| |||||||||||||
Db   135251 GGTCTCACCGTGTAGAGAGGAGGCAGGCCTTGAGGCCACCCCGACACACCAAGAAGGAGAAA 135310

Qy     2520 TCTCGTGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACGAAGGAGGAT 2579
            || |||| || |||| |||||||||| ||||||| ||||||||||||||||||||||||
Db   135311 TTTTGTCGGAAGGACACTCACAGCAGTTCTCGGATCATGTATGTCAACGAAGGAGGAT 135370

Qy     2580 ACCGTCTCCTTGAAGAGAGAAACTCTGTCACTCCTCATGCCTGCTCGTCTAGCTCATCACACCCAT 2639
            |||||||||||||||||||||||| |||||||| ||||||||||||||||||| ||||| ||||||
Db   135371 ACTGTCTCCTTGAAGGGGGAAACTCTGTCACTCCTCATGCCTGCTCGTCTAGGTCATCACACCCCAT 135430

Qy     2640 TTCTCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTC 2699
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135431 TTCTCTTTGCTTCACAGTTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCGTC 135490

Qy     2700 TCTCTCGTTTTATCTCTCCCCTTCTTCTCTCCATCTCCATTCTTTTGA 2759
            |||||       ||| |||||| |||||||||||||||| ||||||||
Db   135491 TCTCTC---TATCTCTCCCTCTTCTCTCTCTCCATTCTCCATTCTTTTGA 135546

Qy     2760 ATTTCCTCATCCCTCCAATCCGTATCTACGCAC-------CCCCCCCCA 2811
            ||||||| ||||||||| || ||||||| ||||       |||||||
Db   135547 ATTTCCCTCATCCCATCCATCCACGCACCCCCCTGCACCCCCCCTCTGCCA 135606

Qy     2812 GGGCAAAGCAGTGCTCTGAGTATCACATCACACAGAAAGGAACAAAAGCGAAACACACAAAC 2871
            |||||||||||||||||||||||||||| |||||||||||||||||||||||| |||||||
Db   135607 GGCAAAGCAGTGCTCTGAGTATCACATCACACAGAAAGGAACAAAAGCAAAACACACAAAC 135666

Qy     2872 CAGCTCAACTTACACTTGGTACTTGGTTACTGGTTACTCAATGGTACTTGTCCTAGCG 2931
            |||||||||||||||||||| |||||||||||||||||||||||||||||||||| 
Db   135667 CAGCTCAACTTACACTTGGTACTTGGTTACTCAATGGTCAATGGTACTTGTCCTAGCA 135726

Qy     2932 TTTTGGAAGAGGAAAACAGGAACCCCTTCCAACCAATCAACCAAACAAAGAAAAAT 2991
            |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   135727 TTTTGGAAGAGGAAAAACAGGAACCTATCAAAACCAATCAAACCAAACAAAGAAAAAT 135786

Qy     2992 TCCACAATGAAAGAATGGTATTTGTCTTTTAAAAAAAAATGTGGAAGTAGAGAAATTTA-CCAAG 3051
            ||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||| ||||
Db   135787 TCCACAATGAAAGAATGATGTTTTTGTCTTTTTT-AAAAAAAAATGTGGAAGTAGAAATTTATACAAG 135846

Qy     3052 AGCAAAATGATTCCTTCTTAAAAATGTGGAAGGGC 3110
            ||||||||||||||||||| |||||||||||||||
Db   135847 AGCAAAATAATTTCTTTCTTT-AAAAAAAAATGTGGAAGGGC 135905

Qy     3111 GTTGTTGGCCCAGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAG-AAGA 3169
            |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   135906 GTTGTTGACCCAGGCGTTAAACTTACAGATTATTTTAACGAGAAAAACACACAGAAAAA 135965

Qy     3170 AGTACCTCAGGTG-TTTTTACCTGACACTTGCTCTTTGTCTTCTGTCTTAGAGATTTTG 3228
            ||||||||||||| |||||||| ||||||||||| ||| ||||||||||||||||||||
Db   135966 AGTACCTCAGGTGTTTTTTTACCTGACACTTGCTCTCTGTGTTCTCTGTTTACAGATTTTG 136025

Qy     3229 TAAAGCTGATAGTTGGAGCATTTTTATTTTTTAAAAATGGAAAAAAAT 3288
            |||||||||||||||| ||||||||||| ||  ||||||||||||||
Db   136026 TAAAACTGACAGTTGGAGCATTTTTAATAATTTTTAATAAAAAGGAGGTTGAAAAAT 136085

Qy     3289 AAGATATCAACTGCCAGCCTGAGAAGAAGTCGACAGTCGCAACAGCTGTTCTGAA 3348
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   136086 AAGATATCAACTGCCAGCCTGAGAAGAAGTCGACAGTCCAAATGTGCAACAGCTGTTCTGAA 136145

Qy     3349 TTGTCTTCCGGCTAGCCAAGAACCNATATGGCCTTCTTTGGACAAACCTTGAAAATGTTT 3408
            ||||||||||||||||||||||||| ||||||||||||||||||| |||||||||||||
Db   136146 TTGTCTTCTGCTAGCCAAGAACCTTATATGACCTTCTTTTGGACAAACTTTGAAAATGTTT 136205

Qy     3409 A 3409
            |
Db   136206 A 136206
```

RESULT 10

```
LOCUS       AL596177            91084 bp    DNA     linear   PRI 08-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-325P15 on chromosome 1, complete
            sequence.
ACCESSION   AL596177
VERSION     AL596177.4  GI:15552942
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 91084)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2001 this sequence version replaced gi:15022347.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep/ This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-325P15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-325P15 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-325P15 is at 1 in this sequence.
            The true left end of clone RP11-337C18 is at 89085 in this
            sequence. The true right end of clone RP4-704D21 is at 41034 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..91084
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-325P15"
                     /clone_lib="RPCI-11.2"
ORIGIN
```

Query Match   77.5%;   Score 2644;   DB 9;   Length 91084;
Best Local Similarity   90.0%;   Pred. No. 0;
Matches 3090;   Conservative   0;   Mismatches 166;   Indels 177;   Gaps   17;

```
Qy      1 GGCAGCGGAGGAGGCGAGGAGCGCGGTACCGGCGCCGGGGGAGCCCGGGGCTCTCGGGG   60
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   9381 GGCAGCGGAGGAGGCGAGGAGCGCGGTACCGGCGCCGGGGGAGCGGCGGGCTCCCGGGG 9440

Qy     61 AAGAGACCGATGATGAACAAGCTTTACATCGGGAACCTGGAACTCGGGAACCTGCATCT  120
          |||||||||||   ||||||| | ||||||||  |||||||
Db   9441 AAGAGACCGGTGATGAACAATCTTTACATCGGGAACAATCTTTACATCGGGAACTTAGCCCTTAGCCCTGGGGAACCTTAGCCCTGGAACGCCGGGTCACCGCCGAC 9500
```

```
Qy    121 GACCTCCGGCAGCTCTTGTGGGGACAGGAGAGTGCCCCTGGCGGGGACAGGGTCTGTGAAG  180
          ||||||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||
Db   9501 GACCTCTGGCAGCTCTTTGGGGACAGGAGAGTGCCCCCACCGGACAGGTCTGTGCTCAAG 9560

Qy    181 TCCGGCTACGCCTTCGTGGGACTACCCGACCAGAGCTGGCGGCCATCCGGCCATCGAGACC  240
          ||||||||||| ||||||||||| |  |||||||||||||||||||||||||||||||||
Db   9561 TCCGGCTATGCCTTCGTGGGACTACCCCTACAAGAACTGGGCCATCCGGCCATCGAGACC 9620

Qy    241 CTCTCGGGTAAAGTGGAATTGCATGGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9621 CTCTCGGGTAAAGTGGAATTGCATGGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 9680

Qy    301 AAGCTAAGGAGGAGAAAATCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGGAGGGTG  360
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   9681 AAGCTAAGGAGGAGAAAATCAGATTCGAAACATCCCTCCTCACCTCCAGTGGGGAGGGTG 9740

Qy    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGACGAATGTGG---AACAAGTCAACACA  417
          |||||||||||||||||||||||||||||||||||||||         |||||||||||
Db   9741 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAACAACAAGTCAACACA 9800

Qy    418 GACACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGAAGCAAAAATAGCC  477
          ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db   9801 GACACAGAAACCACCTGTGTCAACGTCACATGCAACAAGAGAAG-----AAAATAGAC 9855

Qy    478 ATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCG  537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9856 ATGGAGAAGCTAAGCGGGCATCAGTTTGAGAACTAGTCCTTCAAGATTCCTACATCCCG 9915

Qy    538 GATGAAGAGGTGAGCTCCCCTTGCCCCTCAGCGGAGCCCAGCGGTGGGGACCACTCTTCC  597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9916 GATGAAGAGGTGAGCTCCCCTTGCCCCTCAGCGGAGCCCAGCGGTGGGGACCACTCTTCC 9975

Qy    598 C-GGAGCAAGCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATGATTCCCGCT  656
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9976 CGGGGAGCAAGCCACGACCACCCTGGGGCGCTTCTTAGGCCAGACAGATTGATTCGTGCT 10035

Qy    657 GCGGATCCTGGTGCCCCACC-----CAGTTGTGTGCCATCATCGGGAAGAGGAGGGCTT  710
          |||||||||||||||||||          ||||||||||||||||||||||||||||||
Db  10036 GCAGGTTCCTGCTGGCCCTTACAGTTTGCAGTTTGTTGGTTGCCATCATCAGAAAGAGGAGGGGCTT 10095

Qy    711 GACCATAAAGAACAATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAA  770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10096 GACCATAAAGAACATCACTAAGTAGCCCAGTCCCGGATAGATATCCATAGAAAAGAGAA 10155

Qy    771 CTCTGAGCTGCGAGAGAAGCCTGTCACCATCCACCCCAGAGGGGACTTCTGAAGC  830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10156 CTCTGAGCTGCGAGAGAAGCCTGTCACCATCCACCCCAGAGGGGGACTTCTGAAGC 10215

Qy    831 ATGCCCATGATGTTCTTGAAAATCATGCAGAAAGAGAGCAGCAGATGAGACCCAAACTAGCCGAAGA  890
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10216 ATGCCTCATGATTCTTGAAAATCATGCAGAAAGAGAGCAGATGAGCAGAACAAACTAGCCGAAGA 10275

Qy    891 GATTCCTCTGAAAATCTTGGCACACAAATGGCTTGGTTGAAGACTGATTGGAAAAGAAGG  950
          |||||||||||||||||||||                       ||||||||||||||||
Db  10276 GATTCCTCTGAAAATCTTGGCCCACACATGGCTTGG--------TTGGAAAAGAAGG 10323

Qy    951 CAGAAATTTGAAGAAAATTGAACATGAAACATGAAACGCCGGGACCAAGATAACAATCATCTTTGCA  1010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10324 CAGAAATTTGAAGAAAATTGAACATGAAACATGAAACGGGACCAAGATAGCAATCTCATCTTTTGCA 10383

Qy   1011 GGATTTGAGCATTACAACCCGGAAGAACCATCACTGTGAAGGGCACAG-TTGAGGGCCT  1069
          |||||||||||||| |||||||||||||||||||||||||||||||   ||||||||||
Db  10384 GGATTTGAGCATAATAACCAGGAAGAAATCATCACTGTGAAGGGCACAGTTTGAGGGCCT 10443

Qy   1070 GTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAATGATATGC  1129
          |||||||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db  10444 GTGCCAGTGCTGAGATAGAGATTATTGAAGAAGCTGCCTGAGACCTTTGAAAATGATATGT 10503

Qy   1130 TGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTGGGTTGAACCTCAGCGCCACTTGGCATCT  1189
          ||||||||||
Db  10504 TGGCTGTTAA------------------------------------------------------ 10513

Qy   1190 TTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGCGCCCGGGAGTGTCCCCCGCTG  1249
```

```
Right column:

Db  11513 ACATATTAGAAGGGAAAGATGTTAAGATATATATGGCCTGTGGTTACACAGGTGCGCTGCA  11572
           |||||||||||||||||||||||||||||||||||||||||| ||| |||||||||||||
Qy   2329 GCGGTAATATATATTTAGAAAATAATATATCAAATATCAAATTTTAATCA            2388
           ||||||||||||||| ||||||||||||| ||||||||||||| ||  ||
Db  11573 GCAGTAATAATATATTTAGAAAATAATATATCAAATAACTCATCTA----CAATTTTTAATCA  11628

Db   2389 ATTATTAATTTTTTTTTCTTTTAAAGAGAAGAAGCAGGCTTTTCTAGAACTTAAAGAATAA   2448
Qy  11629 ATCATTAA-TTTTTTTTCTTTTAAAGAGAAGAAGCAGGCTTTTCTAGACTTTAAAAGAATA  11687

Qy   2449 A-----GTCTTTGGGAGTCTCACGGTGTAGAGAGAGCTTTGAGGCCACCCGCACAAAA     2503
Db  11688 AAGTCTTTCTTTGGGAGTCTCACAGTATAGACACAGGAGCTTTGAGGCCACCCACACGAAA  11747

Qy   2504 TTCACCCAGAGGGAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTA   2563
Db  11748 TTCACCCCAGAGGGAAATCTCAAAGGAAGGACAATCACAGCAGTTCTGGATCACCTGTGTA  11807

Db   2564 TGTCAACAGAAGGGATACCGTCTCCTTGAAGAGGAAACTCGTCCACTCCTCATGCCGTCTC  2623
Qy  11808 TGTCAACAGAAGGGATACCATCTCCTTGAAGAAGAAACTCGTCACTCCTCATGCCTGTC  11867

Qy   2624 TAGCTCATCACCCATTTCTCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGC      2683
Db  11868 TAGCTCTTACACCCATTTCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGC   11927

Qy   2684 TATATAATTCTCTGTCTCCTCTCTGTTTATCTCTCCCCTCCCCTTCCCCTTCTTCTCC   2743
Db  11928 TATATAATTATCCATCCTCTCTGTTTATCTCTCCCCTCCCCTCCCCGTCTTCTTC      11987

Qy   2744 ATCTCCATTCTTCTTTGAATTCCTCATCCCTCCATCTCAATCCCTATCTACGCACC-CCC  2802
Db  11988 ATCTCCATTCTTTTGAATTCCCCTCCATCCCTCCATCTCAATCCCTATCTATGCACCTCAC  12047

Qy   2803 CCCCCCCAGGCAAAGGAGTGCTCCTGAGTATCACATCACACACAAAGGAACACAAAGCGAAA  2862
Db  12048 TCCCTCACAGGGAAAGCAGTGCTCTGAGTATCGGATCCTGAGATCCCAACAGAAAGTGAAA  12107

Qy   2863 CACACAAACCAGCTCAACTTGCTTGGTTACTTGCTTGGTTACTTCAAAAGAACAAGAGTCAATGGTACTT  2922
Db  12108 CACACAAACCAACCAACCTCAATTTACACTTGGTTACTTACTTGGTTACTTCAAAAGACAAGAGTCAATGGTACAT  12167

Qy   2923 GTCCTAGCGTTTGAGAGAGAAACAGAAACAACAACAGGAAAACAACCACCATCAACAACAA  2982
Db  12168 GTCCTTAGCGTTTGAGAGAGAAACAGAAACAACAGGAAAACAACCATCAAACAACAACAACAA  12227

Qy   2983 AGAAAAAATTCCAACATGAAAGAAGATGTATTTTGTC-TTTTGTCATTTTGGTGTTATAAGGC  3041
Db  12228 AGAAAAAATTCCACATGAAAGAAGAAGGTGTTTTGTTGTCTTTTGTCATTTTGGTGTATAAAGCC  12287

Qy   3042 ATCAATAATTCAGCAAAATGATTCTTTCCTTAAAAAAAAATGTGGAGGAAGTAGAAA      3101
Db  12288 ATCAATATTCAGCAAAATGATTTCTTTCTTTT---AAAAAAATGTGGAGGAAGTAGAAA   12344

Qy   3102 TTTACCAAGGTTGTTGGCCCAGGACGTTTAAATTCACAGATTTTTTTAACGAGAAAAAACAC  3161
Db  12345 TTTACCAAGGTGTGCTGGCCCAGGGCATTAAATTTACAGATTTTTTTAATGGGAAAAAAACAC  12404

Qy   3162 ACAGAAGGAGCTACCTCAGTG-TTTTTACCTACCACTTGCTTCTTGTTTCCCTTAG       3220
Db  12405 ACAGAAAAAGCTGCCTCAGTGTTTTTTACCTGCCACTTGCTTGCTTGTTCCCTTAG   12464

Qy   3221 AGATTTGTAAAGCTGATAGTTGAGCATTTTTTA-TTTTTTTAATAAAATGAGTTGG      3279
Db  12465 AGATTTTGTAAAACTGATAGTTGAGCATTTTTTAATTTTTTTAATAAAATTAGTTGG   12524

Db   3280 AAAAAAAAATAAGATAATCAACTGCAACTGCCAAGTTCCAAGTGTCAACAGC      3339
Qy  12525 AAAAAAAAATAAGATATCAACTGCAACCTGGAGAGGTGCAACAGTGCAACAGC   12584

Qy   3340 TGTTCTGAATTGTCTTCTGCTAGCCAAGACCNATATGGCCTTCTTTTGGCAAAACCTTG   3399
Db  12585 TGTTCTGAATTGTCTTCTGCTAGCCAAGAACCTATATGGCCTTATATGGCCTTCTTTTGGCAAAACCTTG  12644
```

```
Left column:

Db  10514 ------------------------------------  10513

Qy   1250 CCCCCTACCACCCTTCACTACCCTCCGATACTTCTCCCAGCCTGTACCCCATCACC   1309
           ||||||||||||||||||||||
Db  10514 ------------TACGCACTCCGGATACTTCTCCCAGCTTGTACCCCATCACC

Qy   1310 AGTTTGGCCCGTTCCCGCATCATCACTCTTATCCAGAGCGAGGAGATTGAATCTCTTCA   1369
           ||||||||||||||||||||||||||||||||
Db  10555 AGGTTGGCCCGTTCCCGCATCTTATCACTCTTATCCAGAGCAGGAGGTTGAATCTCTTCA  10614

Db   1370 TCCCAACCCAGGCTGTGTGGGCCGCCATCATCGGGAAGAAAGGGGGCACACATCAAACAGTGG   1429
           |||||||||||||||||||||
Qy  10615 TCCCAACCCAGGCTGTGTGGGCCGCCATTATCAGGAGGAGGGAGCACACATCAAACAGCGG  10674

Qy   1430 CGAGATTCGGCCGGAGCCCTCATCAAGATGCCCCTGCGGAAGGCCAGAGACGTCAGCGAAA   1489
           ||||||||||||||||||||||
Db  10675 CGAGATTCGCCACCAGCCCTCCATCAAGATCGCCCTGCGGAAGGGCCCAGACGTCAACGAAA  10734

Qy   1490 GGATGGTCATCATCACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTG   1549
           ||||||||||||||||||||||
Db  10735 GGATGGTCATCATCACCACCCGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTG  10794

Db   1550 GGAAACTGAAAGAGAGAAGAGAAACTTCTTAACCCCAAAGAGAGAAGTGAAGCTGGAGCGCATA   1609
           ||||||||||||||||||||||
Qy  10795 GGAAACTGAAAGAGAGAAGAGAAACTTCTTTAACCCCAAAGAGAAGTGAAGCTGGAGCCCGTA  10854

Qy   1610 TCAGAGTGCCCTCTTCCACAGCTTGCCGTGGCGGTGATTGGCAAGGTGGCAAGACCGTGAACG   1669
           ||||||||||||||||||||||
Db  10855 TCAGAGTGCCCTCTTCCACAGCTGGCCGTGGCGGTGATTGGCAAGGTGTCAATACCTTGAATG  10914

Qy   1670 AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCGAGTGAAA   1729
           ||||||||||||||||||||||
Db  10915 AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACAGGGCCGAGTGAAA  10974

Db   1730 ATGAGGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCGCA   1789
           ||||||||||||||||||||||
Qy  10975 ATGAGGAAGTGATCGTCAGAATTATTGGACACTTCTTTGCTAGCCCAACTGCATAGCACA  11034

Qy   1790 AGATCAGGAGAAATGTACAAACAGGTGAAGCGAGCAGGAGCAGAAATACCCTCAGGGAGTCG   1849
           ||||||||||||||||||||||
Db  11035 AGATCAGGAGAAACTGTACACAACAGGTGAAGCGAGCAGGAGCAGAAATACCCTCAGGGGAGTCG  11094

Qy   1850 CCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAGAACAACGCGGATGAATGTAGC   1909
           ||||||||||||||||||||||
Db  11095 CCTTACAGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAGAACAACGCGGATGAATGGAGC  11154

Db   1910 CCTTCCAACACCTGGCGGGGCGGGGCGGGGCCGGAAGGTCAGCCAGGTCAGCCAGGGCCC   1969
           ||||||||||||||||||||||
Qy  11155 CCTTCCAACACCTGGCGGGGCGGGGCGGGGCCGGAAGGTTCAGCCAGGTCAGCCAGGGCCCC  11214

Qy   1970 GACCATCTGAGGAATGAGAGTCTGGGAGGCGGCGGAACTCTGCCGAGGCCCTGAGA      2029
           ||||||||||||||||||||||
Db  11215 GACCATCTGAGGAATGAGAGTCTGGGAGGCGGCGGGAACTCTGCCGAGGCCCTGAGA   11274

Db   2030 ACCCCAGGGGCCGAGAGGGGCCGGGGCGGGGAAGGGTGGGGAAGGTTGGCCAGAGACCACCGAGCCC   2089
           ||||||||||||||||||||||
Qy  11275 ATTTCCAGGGGCCGAGAGGGGCGGGGAAGGGTGGGGAAGGTTGGCCAGAACCACCGGGCCCC  11334

Qy   2090 -CGCCTCCCGCCCCCCCCCCCCCCCCATCAGGGCTTCTTCTGCAGGCTTCAGCGGCTTCAGCTTCTCAGGAGACTCG   2148
           ||||||||||||||||||||||
Db  11335 GCTTCTCACCCCCATCAGGGCTTCTGCAGGCTTCAGCGGCTTCAGCTTCTCAGGAGACTCCG  11394

Qy   2149 ATCTCTCCTGAACTCCCAGACGGCCATCCCTTTAGTTGAACTAACATAGGTGAACGTGT   2208
           ||||||||||||||||||||||
Db  11395 ATCTCTCCTTAACTCCCAGACGGCCATCCCCTTTAGTTGAACTAACAATAGG--TGATGCGT  11452

Db   2209 TCAAAGCCAAGCAAAATGCACACCCTTTTTCTGTGTGCAAATCGTCTCTGTACATGTGTGT   2268
           ||||||||||||||||||||||
Qy  11453 TCAAAGCCAAGCAAAATGCACACCCCTTTTTTGTGACAAGTCATCTCTGTACATGTGTAT   11512

Qy   2269 ACATATTAGAAAGGGAAGATGTTAAGATATATGTGGCCTGTGGGTTACACAGGTGCGCTGCA   2328
           ||||||||||||||||||||||
Db  11513
```

```
Qy      3400  AAAATGTTTATTT 3412
              |||||||||||||
Db     12645  AAAATGTTTATTT 12657

RESULT 11
AC015706
LOCUS       AC015706              182695 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-325P15 map 1, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC015706
VERSION     AC015706.3  GI:8096833
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182695)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 1, clone RP11-325P15
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 182695)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 182695)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 26, 2000 this sequence version replaced gi:6454072.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1382
                Center clone name: 325_P15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 141788 bases at least Q40
                Consensus quality: 163541 bases at least Q30
                Consensus quality: 174666 bases at least Q20
                Insert size: 181000; agarose-fp
                Insert size: 181095; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     2035:  contig of 2035 bp in length
            *   2036     2135:  gap of 100 bp
            *   2136     6069:  contig of 3934 bp in length
            *   6070     6169:  gap of 100 bp
            *   6170    10460:  contig of 4291 bp in length
            *  10461    10560:  gap of 100 bp
            *  10561    14300:  contig of 3740 bp in length
            *  14301    14400:  gap of 100 bp
            *  14401    18480:  contig of 4080 bp in length
            *  18481    18580:  gap of 100 bp
            *  18581    23093:  contig of 4513 bp in length
            *  23094    23193:  gap of 100 bp
            *  23194    28213:  contig of 5020 bp in length
            *  28214    28313:  gap of 100 bp
            *  28314    32808:  contig of 4495 bp in length
            *  32809    32908:  gap of 100 bp
            *  32909    37000:  contig of 4092 bp in length
            *  37001    37100:  gap of 100 bp
            *  37101    45183:  contig of 8083 bp in length
            *  45184    45283:  gap of 100 bp
            *  45284    53421:  contig of 8138 bp in length
            *  53422    53521:  gap of 100 bp
            *  53522    64476:  contig of 10955 bp in length
            *  64477    64576:  gap of 100 bp
            *  64577    75144:  contig of 10568 bp in length
            *  75145    75244:  gap of 100 bp
            *  75245    93230:  contig of 17986 bp in length
            *  93231    93330:  gap of 100 bp
            *  93331   116951:  contig of 23621 bp in length
            * 116952   117051:  gap of 100 bp
            * 117052   143611:  contig of 26560 bp in length
            * 143612   143711:  gap of 100 bp
            * 143712   182695:  contig of 38984 bp in length
FEATURES             Location/Qualifiers
     source          1..182695
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-325P15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2035
                     /note="assembly_fragment"
     misc_feature    2136..6069
                     /note="assembly_fragment"
     misc_feature    6170..10460
                     /note="assembly_fragment"
     misc_feature    10561..14300
```

```
misc_feature    /note="assembly_fragment"
                14401..18480
misc_feature    /note="assembly_fragment"
                18581..23093
misc_feature    /note="assembly_fragment"
                23194..28213
misc_feature    /note="assembly_fragment
                28314..32808
                clone_end:T7
                vector_side:right"
misc_feature    /note="assembly_fragment"
                32909..37000
misc_feature    /note="assembly_fragment"
                37101..45183
misc_feature    /note="assembly_fragment"
                45284..53421
misc_feature    /note="assembly_fragment
                53522..64476
                clone_end:SP6
                vector_side:left"
misc_feature    /note="assembly_fragment"
                64577..75144
misc_feature    /note="assembly_fragment"
                75245..93230
misc_feature    /note="assembly_fragment"
                93331..116951
misc_feature    /note="assembly_fragment"
                117052..143611
misc_feature    /note="assembly_fragment"
                143712..182695

ORIGIN

Query Match       77.4%;    Score 2642.4;   DB 2;  Length 182695;
Best Local Similarity 90.0%;  Pred. No. 0;
Matches 3089;  Conservative 0;  Mismatches 167;  Indels 177;  Gaps 17;
```

(DNA sequence alignment between query "Qy" and database "Db" sequences)

```
Db  20720 ||||||||||||||||||||||||||||||||||||||||||||||||  |||| |||||
           TCAGAGTGCCCTCCTTCCACAGCTGGCCGGGTGATTGACCAAAGGTGTCAATACCTTGAATG  20779

Qy   1670 AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAGCGCAGATGAAA  1729
          ||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db  20780 AACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAGCCAGATGAAA  20839

Qy   1730 ATGAGGAAGTGATCGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTCGCACAGCGCA  1789
          ||||||||||||||||||| ||||||||| |||| ||||||  ||||| |||||||||||
Db  20840 ATGAGGAAGTGATCGTCAGAATTATTGGACACTTCTTTGCTACCCCAACTGCATAGCACA  20899

Qy   1790 AGATCAGGGAAATTGTACAACAGTGAAGCAGCAGAGGACAGAAATACCCTCAGGGAGTCG  1849
          ||||||||||||||| |||||||||||||| |||||||||| ||||||||||||||||||
Db  20900 AGATCAGGGAAACTGTACAACAGGTGAAGCAGCAGAGCAGAAATACCCTCAGGGAGTCG  20959

Qy   1850 CCTCCACAGCGCCAGCAAGCAAGCGTCCCACAGGCACCAGCAGCACGATGAATGTAGC  1909
          ||||||||| |||||| |||||||||||||||||||||||| |||| ||||||||||||
Db  20960 CCTCACAGCAGCACAGCAGCAAGCGTCCCACAGGCACCAGCAGCAACCAACAATGGAGC  21019

Qy   1910 CCTTCCAACACCTGACAGAATGAGACCAAACGCAGCGAGCCAGATCGGGAGCAAACCAAA  1969
          |||| |||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  21020 CCTTCCAACACCTGACAGAATGAGACCAAACAGCCAGATCGGGAGCAAACCAAA  21079

Qy   1970 GACCATCTGAGGGATGAGAAGTCTGCGAGGCGGCCACAGGGACTCTGCCGAGGCCCTGAGA  2029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21080 GACCATCTGAGGGATGAGAAGTCTGCGGAGAGTCGGGGACTGAGAGGGCCCTGAGA  21139

Qy   2030 ACCCCAGGGGCCGAGGAAGTGGGGGAAGTCAGCCAGGTTGCCAGAACACCGAGCCC  2089
          |||| ||| ||||||||||||||||||| ||||||||||||||||||||| |||||
Db  21140 ATTCCAGGGGCCGAGGAAGTGGGGGAAGTCAGCCAGGTTGCCAGAACCACTGGGGCCC  21199

Qy   2090 -CGCCTCCCGCCCCCCCCCCCAGGGCTTCTGCAGGCTTCAGCCTTCCCACTTCACCACTCGG  2148
          | | ||||||||            ||||||||||||||||||||||||||||||||||||
Db  21200 GCTTCTCACCCCATCAGGGCTTCTGCAGGCTTCAGCCTTCCCACTTCACCACTCCG  21259

Qy   2149 ATCTCTCCTGAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACTAGGTGAACGTGT  2208
          |||||||| ||||||||||||| ||||||||||||| |||| ||||||| || |||||
Db  21260 ATCTCTCCTTAACTCCCACGACGCTATCCCTTTTAGTTGAACTAACTAAGTAG--TGATGCGT  21317

Qy   2209 TCAAAGCCAAGCAAAATGCACACCCTTTTCTGTGGCAAATGCTCTCGTACATGTGTGT  2268
          |||||||||||||||||||| |||| |||||||||| |||| | |||||||||||||||
Db  21318 TCAAAGCCAAGCAAAATGCACACACCCTTTTTGTGACAAGTCATCTCGTACATGTGTAT  21377

Qy   2269 ACATATTAGAAAGGGAAGATGTTAAGATATGTGGCCTTGTGGGTTACACAGGGTGCCTGCA  2328
          |||||||| |||||||||||||| |||||||||||||||||||||||||||||||||||
Db  21378 ACATATTAGAAGGGAAAGATGTTAAGATATGTGGCCTTGTGGGTTACACAGGGTGCCTGCA  21437

Qy   2329 GCGGTAATATATTTAGAAATAATATCAAATAACTCAACTAACTCCAATTTTAATCA  2388
          || ||||||||||||||||| |||||||||||||||||||||||      ||||||||
Db  21438 GCAGTAATATATTTAGAAATAATATCAAATAACTCAACTAACTCATCTA----CCATTTTAATCA  21493

Qy   2389 ATTATTAATTTTTTTTTCTTTTTCTTTTTCTTTTTAAAGAGAAAAGCAGGCTTTTCTTTT  2448
          || |||| |||||  |||||||||| ||||||||||||||||||||||||||||||||||
Db  21494 ATCATTAA-TTTTTTTTCTTTTTCTTTTTAAAGAGAAAAGCAGGCTTTTCTTTT  21552

Qy   2449 A-----GTCTTTGGGAGGTCTCACGGGTGTAGAGAGAAGCTTTGAGGCCACCCGCACAAAA  2503
          |     ||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  21553 AAGTCTTTCTTTGGGAGGTCTCACAGTATAGACAGAGGCTTTGAGGCCACCCACACAAAA  21612

Qy   2504 TTCACCCGAGGGAAATCTCGTCGGAAGGACACTCAGGGCAGTTCTCGGATCACCGTGTA  2563
          ||||||| |||||||| |||||||||||| |||||||||| |||||||||||||||||||
Db  21613 TTCACCCAGAGGGAAATCTCAAAGGAACGGACAATCACACAGCAGTTCTGGATCACCGTGTA  21672

Qy   2564 TGTCAACAGAGGGATACCGTCCCTTGAAGAGGAAACTCTGCACTCCTCATGCCTGTC  2623
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21673 TGTCAACAGAGGGATACCATCCCTTGAAGAGGAAACTCTGCACTCCTCATGCCTGTC  21732

Qy   2624 TAGCTCATACACCCATTTCTTTGCTTCACAGGTTTTTTGCATACTGC  2683
          |||||  ||||| |||||||||||||| |||||||| ||||||||||
Db  21733 TAGCTCTTTACACCCCATTTCTTTGCTTCACAGGTTTTTTAAACTGGATTTTTGCATACTGC  21792

Qy   2684 TATATAATTCTCTGTCTCTCGTTTATCTCGTTTCCCCTTTCTTCTCC  2743
          ||||||  |||| ||||||||||||| ||||||||||||||||||||
```

```
Db  21793 TATATAATTATCCATCTCTCTGTTTATCTCCCTCCCCATCTCCCGTTCTTCTTC  21852

Qy   2744 ATCTCCATTCTTTTGAATTCCTCATCCCTCCATCTCAATCCCGTATCCTACGCACC-CCC  2802
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  21853 ATCTCCATTCTTTTGAATTCCCTCATCCCTCCATCTCAATCCCGTATCCTACGCACC  21912

Qy   2803 CCCCCCCAGGCAAAGCAGTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCCGAAA  2862
          |||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  21913 TCCCTCACAGGGAAAGCCAGCAGTGCTCTGAGTATCGCATCACAGGGAACAACAAAGTGAAA  21972

Qy   2863 CACACAAACCAGCCTCAACTTACACTTGGTTACTCAAAGAGACAGAGTCAATGGTACTT  2922
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  21973 CACACAAACCAACCTCAATTTACACTTGGTTACTCAAAGAGAACAAGAAGGTCAATGGTACAT  22032

Qy   2923 GTCCTAGCGTTTTGGAAGAGGAAAACAGGAACCACCAAAACCAATCAACCAACAA  2982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22033 GTCCTAGCGTTTTGGAAGAGGAAAACAGGAACCCATCAAACAACCAACCAACAA  22092

Qy   2983 AGAAAAAATTCCACAATGAAAGAATGTATTTTGTC-TTTTTGCATTTTGGTGTATAAGCC  3041
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  22093 AGAAAAAAATTCCACAATGAAAGAAGAGGTGTTTTGTCTTTTTGCATTTTGGTGTATAAGCC  22152

Qy   3042 ATCAATATTCAGCAAAATGATTCCTTTTAAAAAAAATGTGGAGGAAAGTAGAAA  3101
          |||||||||||||||||||||||| ||||||||   |||||||||||||||||||
Db  22153 ATCAATATTCAGCAAAATGATTTCTTTCTTT---AAAAAAAATGTGGAGGAAAGTAGAAA  22209

Qy   3102 TTTACCAAGGTTGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACAC  3161
          |||||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db  22210 TTTACCAAGGTTGCTGCCCAGGGCATTAAATTTACAGATTTTTTTAATGGGAAAAACAC  22269

Qy   3162 ACAGAAGAAGTCACTACCTCAGGTG-TTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAG  3220
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  22270 ACAGAAAAAGTGCCTCCAGGTGTTTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAG  22329

Qy   3221 AGATTTTGTAAAGCTGATAGTTGGAGCATTTTTTA-TTTTTTTAATAAAAATGAGTTGG  3279
          |||||  ||||||||| ||||||||||||||||||| |||||||||||||||||||||||
Db  22330 AGATTTTGTAAAACTGATAGTTGGAGCATTTTTTTAATTTTTTTTAATAAAATTAGTTGG  22389

Qy   3280 AAAAAAAATAAGATATCAACTGCCAGCCTGGAGAGGTGACAGTCCAAGTGTGCAACAGC  3339
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  22390 AAAAAAAAATAAGATATCAACTGCCAGCCTGGAGAGGTGACAGTGCAACAGC  22449

Qy   3340 TGTTCTGAATTGCTCTCCGCTAGCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTG  3399
          |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db  22450 TGTTCTGAATTGTCTTCTGCTAGCCAAGAACCTATATGGCCTTCTTCTTTGGACAAACCTTG  22509

Qy   3400 AAAATGTTTATTT  3412
          |||||||||||||
Db  22510 AAAATGTTTATTT  22522
```

RESULT 12
LOCUS   AC104980   187226 bp   DNA   linear   PRI 31-MAR-2002
DEFINITION   Homo sapiens chromosome 8, clone RP11-281D17, complete sequence.
ACCESSION   AC104980
VERSION   AC104980.5   GI:19849375
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 187226)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 8, clone RP11-281D17
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 187226)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
             Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
             Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

```
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 187226)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkly,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 187226)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkly,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Mar 31, 2002 this sequence version replaced gi:19683374.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------- Project Information
                Center project name: L2878
                Center clone name: 281_D_17
            --------
FEATURES             Location/Qualifiers
     source          1..187226
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-281D17"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1..5915
                     /rpt_family="L1PA5"
     repeat_region   5916..5943
                     /rpt_family="AT_rich"
     repeat_region   5944..6477
                     /rpt_family="L1"
     repeat_region   8563..8886
                     /rpt_family="MER2"
     repeat_region   complement(9296..9555)
                     /rpt_family="L2"
     repeat_region   complement(9865..10094)
                     /rpt_family="MIR"
     repeat_region   complement(10346..10504)
                     /rpt_family="AluSg/x"
     repeat_region   complement(10821..11111)
                     /rpt_family="AluSc"
     repeat_region   11162..11300
                     /rpt_family="MER47A"
     repeat_region   11301..11596
                     /rpt_family="AluSx"
     repeat_region   11597..11836
                     /rpt_family="MER47A"
     repeat_region   12185..12319
                     /rpt_family="FLAM_C"
     repeat_region   12520..12822
                     /rpt_family="AluSp"
     repeat_region   complement(12885..13167)
                     /rpt_family="AluJb"
     repeat_region   complement(13230..13379)
                     /rpt_family="MIR"
     repeat_region   complement(14179..14479)
                     /rpt_family="AluSg"
     repeat_region   complement(14630..14747)
                     /rpt_family="MIR"
     repeat_region   14748..15055
                     /rpt_family="AluSq"
     repeat_region   complement(15056..15145)
                     /rpt_family="MIR"
     repeat_region   complement(15206..15400)
                     /rpt_family="MIR"
     repeat_region   15576..15606
                     /rpt_family="AT_rich"
     repeat_region   complement(15607..15901)
                     /rpt_family="AluSp"
     repeat_region   16127..16556
                     /rpt_family="L1ME4A"
     repeat_region   16657..16684
                     /rpt_family="AT_rich"
     repeat_region   complement(17149..17513)
                     /rpt_family="MLT1A1"
     repeat_region   17966..18067
                     /rpt_family="MER47B"
     repeat_region   complement(18068..18365)
```

```
repeat_region    /rpt_family="AluJo"
                 18366..18666
repeat_region    /rpt_family="MER47B"
                 19538..19561
repeat_region    /rpt_family="AT_rich"
                 complement(19562..21070)
repeat_region    /rpt_family="L1PA2"
                 complement(21072..21141)
repeat_region    /rpt_family="U3"
                 22731..22771
repeat_region    /rpt_family="MIR"
                 complement(23445..23902)
repeat_region    /rpt_family="L1PA5"
                 complement(25000..25014)
repeat_region    /rpt_family="FLAM_C"
                 25015..25083
repeat_region    /rpt_family="(TTCA)n"
                 complement(25084..25200)
repeat_region    /rpt_family="FLAM_C"
                 25393..25450
repeat_region    /rpt_family="AT_rich"
                 complement(25762..26547)
repeat_region    /rpt_family="L1ME4A"
                 26829..26889
repeat_region    /rpt_family="(TAGA)n"
                 27373..27795
repeat_region    /rpt_family="LTR37A"
                 complement(29177..29247)
repeat_region    /rpt_family="MIR"
                 29248..29370
repeat_region    /rpt_family="AluJo"
                 29371..29395
repeat_region    /rpt_family="(CA)n"
                 29396..29562
```

```
Query Match          64.2%;  Score 2189.6;  DB 9;  Length 187226;
Best Local Similarity 85.8%;  Pred. No. 0;
Matches 2727;  Conservative  0;  Mismatches  210;  Indels  242;  Gaps  17;

Qy    247  GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTTCTAAAAGCTA  306
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  79640  GGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCGGTTCTAAAAGCTA  79699

Qy    307  AGGAGCAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTGTTGGAT  366
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79700  AGGAACAGAAAATTCAGATTTGAAACATACCCTCCTCACCTGCAGTGGGCGGATGTTGGAT  79759

Qy    367  GGACTTTTGGCTCCAATATGGGACAGTGGAGAGATGTGAGAAGTGAACAAGAACACACAGAA  426
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  79760  GCACCCTTTGGCTCAACACGGGATAGTAGTGAGAATGTGAACAAGAACAAGAACACACAGAA  79819

Qy    427  ACCGCCGTTGTCAACGTCACATATGCAACAAGGAAGAAGCAAAAAATAGCCATGGAGAAGAAG  486
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  79820  ACTGCTGTTGTCAATGTCACATATGCAACAAGGGAAGAAGC-AAAATAGCCATGGAGAAGAAG  79878

Qy    487  CTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTTCACATCCCTACATCCCGATGAAGAG  546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79879  CTAAGCGGGCATCAGTTTGAGAACTATTCCTTCAAGATTCCTTCACATCCTGGATGAAGAG  79938

Qy    547  GTGAGCT-CCCCTTCGCCCCCTCGGCCCCTCGGCAGCCCAGCCCAGCCCGGCCACTCTTCCCGGGAGCA  605
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79939  GTGAGCTCCCCCTTACCCCCTTCGGCCCCCTCGGCAGCCCAGCCCGGGGACCCACCACTCTT-CCGGGAGCA  79997

Qy    606  AGGCCACGCCCCTGGGGGCACTTCTCAGGCCCAGACAGATTGATTTCCCGCTGCCGGATCCT  665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79998  AGGTCACC-CCCTGGGGGCCTCTTCTCAGGCCCGGACACGATTTCCCGCTGCCGGATCCA  80056

Qy    666  GGTCCCACCCAGTTGTTGTGGTGCCATCATCCGGAAAGGAGGGCTTGACCATAAAGAACAT  725
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  80057  ------ACCCAGTTTGTTGTGCCATGGTCTGAAAGGAGGGCTTGAACATTTGAAAGTGAACAT  80109

Qy    726  CACTAAGCAGACCCAGTCCCGGGTAGTATCCATAGAAAGAGAACTCTGGAGCTGCAGA  785
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80110  CACTGAGCAGACCCAGTCCCGGGTAGACATCCATAGAAAGAGAACTCTGGAGCTGCAGA  80169

Qy    786  GAAGCCTGTCACCATCCATGCCACCCCAGAGGGAGCTTCTGAAGCATGCCGCATGAGATTCT  845
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80170  GAAGCCTTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCACCACGATTCT  80229

Qy    846  TGAAATCATGCAGAAAGCAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAAT  905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80230  TGAAAGCATGCAGAAAGCAGGCAGATGAGACCAAACTAGCCCGAAGAGATTCCTCTGAAAAT  80289

Qy    906  CTTGGCACACAATGGCTGGTTGGAAGACTGATTGGAAGAAAGAAGGCAGAAATTTGAAGAA  965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80290  CGTGGCCCACAATGGCTTGTTGTTTGGAAGACTGATTGGAAGAAAGAAGGCAGAAATGTGAAGAA  80349

Qy    966  AATTGAACATGAAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATA  1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80350  AAATGAACATGAAACAGGGACCAAGATAACAATCTCATCCTCCTGCAGGATTTGAGCATATG  80409

Qy   1026  CAACCCGGAAAGAACCATCACTGTGGAGGGCACAGTTGAGCCTGTGCCAGTGCTGAGAT  1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80410  CAACCCGGAAAGAACCATCACTGTGAAGGGCCACAGTTGAGCCTCATGCCAGTGCTGAGAT  80469

Qy   1086  AGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACA  1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||-----
Db  80470  GGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAA------  80523

Qy   1146  AGCCAATCTGATCCCAGGGTTGAACCTGCAGTACATTGGCATCTTTTCAACAGGACTGTC  1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80524  -----------------------------------------------------------  80523

Qy   1206  CGTGCTATCTCCACCAGCAGCAGGGCCCCGCGGAGCTCCCCCGCTGCCCCCTACCACCCCTT  1265
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80524  -----------------------------------------------------------  80523

Qy   1266  CACTACCACTCCGGATAC-TTCTCCAGCCTGCACCCCATCACCAGTTTGGCCCGTTCC  1324
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  80524  ---TACCCACTCCAGATACTTCTTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTCC  80580

Qy   1325  CGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTG  1384
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  80581  CGCATCAGCATTCTTATCAAGAGCAGGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTG  80640

Qy   1385  TGGGGGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGCGAGATTCGCGGGAG  1444
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  80641  TGGGTACCATCATCGGGAAGAAAGGGGCACACATCAAACAGCTGCGAGACTCGCCAGAG  80700

Qy   1445  CCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAGGATGGTCATCATCA  1504
           ||||||||||||||||||||||||||||||
Db  80701  CCTCCA-----------------------------------------------------  80706

Qy   1505  CCGGGCACCGGAAGGCCCAGTTCAAGGCCCAGGAGCGGATCTTTGGGAAACTAAAGAAG  1564
           |||||||||||||||||||||||||||||||||||||||||||
Db  80707  --------------TCAAGGGCCAGGAGCGGATCTTTCCGGGAAACTAAAAGAG  80745

Qy   1565  AAAACTTCTTTAACCCAAGAAGAAGTGGAAGGCATATCAGAGTGCCCTCTT  1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80746  AAAACTTCTTTAACCCAAGAAGAAGTGGAAGCGCCCATATCAGAGTGCTCTCTT  80805

Qy   1625  CCACAGCTGGCCGGGTGATTGGCAAGAGTGGCAAGACCGGCAAGAACTGCAACTTAA  1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80806  CCACAGCTGGCCGGGTGATTGGCAAGAGTGGCAAGACTGGCAAGAACTGCATAACTTAA  80865

Qy   1685  CCAGTGCAGAATGCATCGTGCCTGTGACAGATGAAAATGAGGAAGTGATCG  1744
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80866  CTAGTGCAGAAATCATTGTGCCTGTGACCAAACGCCCGATGAAGAAGTGATCA  80925

Qy   1745  TCAGAATTATCGGGCACTCTTTGTCTAGCCAGACTGCACAGCCCAAGATCAGGGAAATTG  1804
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80926  TCAGAATTATCGGGCACTCTTTGTCTAGCCAGACTGCACAGCCCAGTATCAGGGAAACTG  80985

Qy   1805  TACAACAGGTGAAGCGAGCAGCAGAATACCCTCAGGGAGTCGCCTCACAGCCGCAGCA  1864
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80986  TACAAGAGGTGAAGCGAGCAGCAGAGTAGAAATACCCTCAGGGAGTCGCCTCACAGCCGCAGCA  81045
```

```
Qy   1865  AGTGAGG-CTCCCACAGGCACAGCAAAACACGGATGAATGTAGCCCTTCCAACACCTG  1923
            ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  81046  AGTGAGGCCTCCCACAGGCATCAGCAAAACAACAGATGAATGGAGCCCTTCCAACACCTG  81105

Qy   1924  ACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAGACCATCTGAGGAA  1983
            |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  81106  ACAAAATGAGACCAAACAACAGCCAGCCAGATCGGAGCAAACCAAGACCATCTGAGGAA  81165

Qy   1984  TGAGAAGTCTGCGAGGGCCGGCCGAGGGCCTGAGGAACCCAGGGGCCGA  2043
            |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  81166  TGAGAAGTCTGCGAGGGCCGGCCGAGGGCCCTGAGAACTCCAGGGGCCAA  81225

Qy   2044  GGAGGGGCGGGGAAGGTCAGCCAGGTTGCCAGCAACCACGAGCCCGCCCCC  2103
            ||||  | |||| |||||||||||||||||||||||||||||||||||||||
Db  81226  GGAGGCG-GGGGAGGGTCAGCCAGGTTGCCAGCAACCACCTACCTCGGGGCCCTGCCTACCTCC  81284

Qy   2104  CAGGGCTTCTGCAGGCTTCAGCCATCCACCTCACCTCCTCTGAACTC  2163
            ||||||||||||||| |||||||||||||||||||||||||||||||
Db  81285  CAGGACTTCTGCA-GCTTCAGCAAT CCACCTCACCTCCTTAACTC  81343

Qy   2164  CCAGACGCTATCCCTTTAGTTGAACTAACATAGTGAACGTGTTCAAGCCAAGCAAA  2223
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  81344  CCACAATGCTGTACCTTTTGGTTGAAGTAACATAGGTGAACGTGTTCAAGCCATGCAAA  81403

Qy   2224  ATGCACACCCTTTTCTGTGTGCAAAT---TCTCGTACACGTGTGTACATATTAGAAAGGG  2283
            ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db  81404  ATGCACACCCTTTTTGTGTGCAAA----TCTCGTACATGTGTGCACATATTAGAAGGGG  81459

Qy   2284  AAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTT  2343
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81460  ACGATGTAAGATACGTGGCCTGCGCGGGTGCCTGCAGCGGTAATATGTTTT  81519

Qy   2344  AGAAATAATAATATCAAATAACTCAACTAACTCAATTTTTAATCAATTATTAATTTTTTT  2403
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81520  AGAAATAACAATATCAAATAACTCAACTAACTCAATTTTCCAATTTTAATCAACTATTAATTTTTT  81579

Qy   2404  TTCTTTTTAAAGAGAAAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTC  2463
            |||||||||||||||||||||||||||||||||||      ||||| |||||||||
Db  81580  TTCTTTTTAAAGAGAAAGAAAGCAGGCTTTTCTAGACTTTAAAGTCT---TTCTTTGGGAGGT  81635

Qy   2464  TCACGGTGTGAGAGGAGTTTGAGCGCCACCCGCACAAAATTCACCCAGAGGGAAATCTC  2523
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81636  TCACGGTGTAGAGGAGAGCTTCGAGGTGACCTACACAAAATTCACCCAGAGGGAAATCTT  81695

Qy   2524  GTCGGAAGGACACTCACGGCAGTTCTGGATCACCGTGTATGTCAACAGAGAGGGATACCG  2583
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  81696  GTCGGAAGGACACTCACAGCAGCAATTGTGTATCGTATCTGTGTAATCAACAGAGGGATACCA  81755

Qy   2584  TCTCCCTTGAAGGAGAAACTCTGTCACTCCTCCATGCCTGTCTCCAGCTCATCACCCATTTCT  2643
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  81756  TCTCCCTTGAAAAGGAGAAACTCTGTCACTCGTCTGTCTCCAGCTTATCATCACATCCATTTCT  81815

Qy   2644  CTTTGCTTCACAGGTTTTAAACTGGTTTTTTTCATACTGCTATATAATTCTC--TGTCTC  2701
            |||||||||||||||||||||||||||||||||||||||||| |||||||||   |||||
Db  81816  CTTTGCTTCATGGGTTTTTAAACTGGTTTTTTGCATACTGCTATATAATAATTCTCTATCTCTC  81875

Qy   2702  TCTCTGTTATCTCTCCCCTCCCCCTCCCCCTCCCCCCTCTTTTTGAAT  2761
            |||||| ||||||||||||||||||||||||      ||||||||||||
Db  81876  TCTCTGTTTATCTCTCACCTCCTCACCTCCCCCCCCC-------CCCCGCATCTTTTTGAAT  81935

Qy   2762  TTCCTCATCCCTCCATCTCAATCCCGTATCTACGCA-------CCCCCCCCCCCCAGG  2813
            |||||||||||||||||||||||||||||||||||        |||||||||||||||
Db  81936  TCCCTCATCCCTCCATCTCAATCCCATATCTACGCAACACCACCCTGCGGCCCCCGCCCCAAG  81995

Qy   2814  CAAAGCAGTGCTCTGAGTATCATCATCCAACACACAGGAACAAACCGAAACACACAACCA  2873
            ||||||||||| |||||||||||| ||||||||||||||||||||| ||||||||||||||
Db  81996  GAAAGCAGTGCTTTGAGTATCCAACACACAGGAACAAACAAAAGCAAAAGCACACAACACGTCA  82055

Qy   2874  GCCTCAACTTACACTTGGTTACTCAAAGAAAGGTCAATGGTACTTGTCCTAGCGTT  2933
            |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  82056  GCCTCAACTTACACTTGGTTACTCAAAGGTCAATGGTACTTGTCCTAGCATT  82115
```

```
Qy   2934  TTGGAAGAGGAAAACAGGAACCCACCCACCAACCAATCACCAACCAAACAAAGAAAAATTC  2993
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82116  TTGGAAGAGGAAAACAGGAACCCACCCACCAACCAATCAATCACCAACCAAACAAAGAAAAAATTC  82175

Qy   2994  CACAATGAAAGAATGTATTTTGTCTTTTGCATTTGTGTATAAGCCATCATATTCAG  3053
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82176  CACAATGAAAGAATGTATTTTGTCTTTTGCATTTGTGTATAAGCCATCATATTCAG  82235

Qy   3054  CAAAATGATTCTTTCTTTAAAAAAAAAATGTGGAGGAAAGTAGAAATTACCAAGGTT  3113
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  82236  CAAAATTATTTCTTTCTTTTTTAAAAATGTGAAGGAAGTAGCAATTATACAAGGTT  82295

Qy   3114  GTTGGCCCAGGGGCGTTAAATTCACAGATTTTTTTTACGAGAAAAAACACAGAGAGAGCT  3173
            |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  82296  GTTGGCCCAGGGGCGTTAAATTTACCGATTTTATAAATGAGAAAAAACACAGAGAAAAAGCT  82355

Qy   3174  ACCTCAGGTG-TTTTACCTCAGCACCTTGCTCTTGTCCTTGTTCCCTTAGAGATTTTGTAAA  3232
            |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82356  ACCTCAGGTGTTTTTGCCTCAGCACCTTGCTCTTGTCCTTGTTCCTTTCCTTTAGAGATTTTGTAAA  82415

Qy   3233  GCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAATGAGTTGGAAAAAAATAAGA  3292
            ||||||||||||||||      ||||||||||||||||||||||||||||||||||
Db  82416  ACTGATAGTTGGAGCA----TTTATTTTCTTAATAAAAATGAGTTGGAAAAAAATAAGA  82471

Qy   3293  TATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCGTGTTCGAATTGT  3352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82472  TATCAACAACCAGCCTGGAGAAGGTGAGAGTCCAATTGTGCAACAACTGTTCTGAATTGT  82531

Qy   3353  CTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTTTGACAAACCTTGAAAATGTTTATT  3411
            |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db  82532  CTTCTGCTAGCCAAGAACCTATATGGCCCTTCTTTTTAGACAACCTTGAAAACATTTATT  82590
```

```
RESULT 13
AP004290/c
LOCUS       AP004290                113201 bp    DNA     linear   PRI 04-DEC-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone: KB261B8.
ACCESSION   AP004290
VERSION     AP004290.2  GI:17298202
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Homo sapiens DNA chromosome 8 SEQUENCE
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 113201)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Dec 3, 2001 this sequence version replaced gi:16303299.
FEATURES             Location/Qualifiers
     source          1..113201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB261B8"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   131..298
                     /evidence=not_experimental
                     /rpt_family="CT-rich"
     repeat_region   490..512
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
```

```
repeat_region    1347..1650
                 /evidence=not_experimental
                 /rpt_family="AluSc"
repeat_region    3263..3388
                 /evidence=not_experimental
                 /rpt_family="MIR"
repeat_region    3401..3792
                 /evidence=not_experimental
                 /rpt_family="MLT1J"
repeat_region    3796..4090
                 /evidence=not_experimental
                 /rpt_family="AluJb"
repeat_region    4165..4258
                 /evidence=not_experimental
                 /rpt_family="MIR"
repeat_region    complement(4257..4805)
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    complement(4865..5168)
                 /evidence=not_experimental
                 /rpt_family="L1MD2"
repeat_region    5178..5495
                 /evidence=not_experimental
                 /rpt_family="L1MC/D"
repeat_region    complement(5495..6055)
                 /evidence=not_experimental
                 /rpt_family="AluJo"
repeat_region    complement(6057..6348)
                 /evidence=not_experimental
                 /rpt_family="L1M4"
repeat_region    complement(6475..6762)
                 /evidence=not_experimental
                 /rpt_family="AluJo"
repeat_region    complement(6800..7005)
                 /evidence=not_experimental
                 /rpt_family="MLT1J"
repeat_region    complement(7014..7295)
                 /evidence=not_experimental
                 /rpt_family="AluSx"
repeat_region    complement(7402..7456)
                 /evidence=not_experimental
                 /rpt_family="L1MD2"
repeat_region    7690..7746
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    7993..8287
                 /evidence=not_experimental
                 /rpt_family="AluJo"
repeat_region    8341..8400
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    complement(9219..9514)
                 /evidence=not_experimental
                 /rpt_family="AluSp"
repeat_region    9566..9901
                 /evidence=not_experimental
                 /rpt_family="AluSx"
repeat_region    complement(10118..11352)
                 /evidence=not_experimental
                 /rpt_family="MER1B"
repeat_region    complement(11355..11428)
                 /evidence=not_experimental
                 /rpt_family="L1MC1"
repeat_region    complement(11429..11730)
                 /evidence=not_experimental
                 /rpt_family="L1MEc"
repeat_region    complement(11731..12789)
                 /evidence=not_experimental
                 /rpt_family="AluSp"
repeat_region    12790..13085
                 /evidence=not_experimental
                 /rpt_family="L1MEc"
repeat_region    complement(13086..14349)
                 /evidence=not_experimental
                 /rpt_family="AluJb"

repeat_region    14367..14502
                 /evidence=not_experimental
                 /rpt_family="L1MEc"
repeat_region    14724..14905
                 /evidence=not_experimental
                 /rpt_family="L1ME"
repeat_region    15003..15180
                 /evidence=not_experimental
                 /rpt_family="L1ME"
repeat_region    15217..15433
                 /evidence=not_experimental
                 /rpt_family="L1M4"
repeat_region    16753..16837
                 /evidence=not_experimental
                 /rpt_family="L1ME3"
repeat_region    17096..17145
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    17146..17170
                 /evidence=not_experimental
                 /rpt_family="Tigger1"
repeat_region    17171..18147
                 /evidence=not_experimental
                 /rpt_family="(TG)n"
repeat_region    18148..18427
                 /evidence=not_experimental
                 /rpt_family="Tigger1"
repeat_region    18428..19574
                 /evidence=not_experimental
                 /rpt_family="AluSg"
repeat_region    complement(19575..19852)
                 /evidence=not_experimental
                 /rpt_family="Tigger1"
repeat_region    19853..20035
                 /evidence=not_experimental
                 /rpt_family="AluSg"
repeat_region    20272..20565
                 /evidence=not_experimental
                 /rpt_family="Tigger1"
repeat_region    complement(20853..20937)
                 /evidence=not_experimental
                 /rpt_family="AluSq"
repeat_region    complement(21626..21927)
                 /evidence=not_experimental
                 /rpt_family="MLT1H"
repeat_region    complement(22312..24467)
                 /evidence=not_experimental
                 /rpt_family="AluSx"
repeat_region    complement(24463..24524)
                 /evidence=not_experimental
                 /rpt_family="pTR5"
repeat_region    complement(24523..24823)
                 /evidence=not_experimental
                 /rpt_family="HERV9"
repeat_region    complement(24954..25233)
                 /evidence=not_experimental
                 /rpt_family="Harlequin"
repeat_region    complement(25679..25860)
                 /evidence=not_experimental
                 /rpt_family="L1MA8"
repeat_region    complement(25861..26177)
                 /evidence=not_experimental
                 /rpt_family="MER58D"
repeat_region    complement(26178..26364)
                 /evidence=not_experimental
                 /rpt_family="AluSg"
repeat_region    26399..26798
                 /evidence=not_experimental
                 /rpt_family="MER58D"
repeat_region    27355..27395
                 /evidence=not_experimental
                 /rpt_family="L1"
                 /evidence=not_experimental
```

```
repeat_region      /rpt_family="AT_rich"
                27958..27979
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   29533..29562
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   30365..30397
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   30655..30712
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   30881..30935
                /evidence=not_experimental
                /rpt_family="AT_rich"

Query Match       64.1%;  Score 2188;  DB 9;  Length 113201;
Best Local Similarity  85.8%;  Pred. No. 0;
Matches 2726;  Conservative 0;  Mismatches 211;  Indels 242;  Gaps 17;
```

```
Qy    247 GGTAAAGTGAATTGCATGGGAAAATCATGGAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTA 306
          ||||||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||||||||
Db  94347 GGTAAAGTGAATTGTCATGTGGGAAAATCATGGAAGTTGATTACTCGGTCTCTAAAAAGCTA 94288

Qy    307 AGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCCACCTGACAGTGGGAGGTGTTGGAT 366
          ||||||||||||||||||||||||||||| |||||||||||||||||| ||||||||||
Db  94287 AGGAACAGGAAAATTCAGATTTGAAACATACCTCCTCCACCTGACAGTGGGCGATGTTGGAT 94228

Qy    367 GGACTTTTGGCTCAATATGGACAGTGGAAATGTGGAGAAGTGGAACAAGTCAACACAGACACAGAA 426
          |||||||||||||||||||| ||||||||||||||||| ||||||||||||| ||||||||||
Db  94227 GCACCTTTGGCTCAACACGGATAGTGGAGAATGTGGAACAAGTCAACACAGACACAGAA 94168

Qy    427 ACCGCCGTTGCAACGTCACATATGCAACAAGAGAAGCAAAAATAGCCATGGAGAG 486
          |||||||||||||||||||||| ||||||||||||||||||||
Db  94167 ACTGTCTGTTGTCAATGTCACATATGCAACAAGAGGAAGAAGC-AAAATAGCCATGGAGGAG 94109

Qy    487 CTAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAG 546
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  94108 CTAAGGCGGCATCAGTTTGAGAACTATTCCTTCAAGATTCCTACATCCTGGATGAAGAG 94049

Qy    547 GTGAGGT-CCCTTCGCCCCCCCAGCCTGGGAGCCCAGGCCACCACCTCTTCCCGGGAGCA 605
          |||||||  |||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  94048 GTGAGCTCCCCCTTTACCCCCTCAGCCGAGCCCAGGCTGGGGACCCACTCTT-CCGGGAGCA 93990

Qy    606 AGGCCAGCGCCCCTGGGGGCACTTCAGGGCCAGACAGATTGATTTCCCGCTGCCGGATCCT 665
          ||||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||
Db  93989 AGGTCACG-CCCTGGGGGCTTCTTCTCAGGGCCAGGACCGACGACAGATTTCCCGCTGCCGGATCCA 93931

Qy    666 GGTCCCACCAGTTTGTTGGTGCCATCATCGGAAAGGAGGCTTGACCATAAAGAACAT 725
          ||  ||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  93930 ------ACCCAGTTTGTTGGTGCCATGGTCTGAAAGGAGGCTTGACCATAGTGAACAT 93878

Qy    726 CACTAGCAGACGACCCGGTAGTATCCATAGAAAAGAGACTCTGGAGCTGCAGA 785
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93877 CACTGAGCAGACCAGTCCCGGTAGAGACTAGCCATAGAAAAGAGACTCTGGAGCTGCAGA 93818

Qy    786 GAAGCCTGCACCATCCATGCCACCCCAGAGGGGGACTTCTGAAGCATGCGCCATGATTCT 845
          |||||||||||||||||||||||| |||||||||||||||||||||| |||||||||||
Db  93817 GAAGCTTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCCACCACGATTCT 93758

Qy    846 TGAAAATCATGCAGAAAGAGCAGCAGATGAGATTAGCCGAAGAGAGATTCCTCTGAAAAT 905
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  93757 TGAAAGCATGCAGAAGAGGAGCAGATGAGATTAGCCGAAGAGAGATTCCTCTGAAAAT 93698

Qy    906 CTTGGCACCAATGGCTTGTTGGAAGACTAACAATTCATCTTTGCAGGATTTGAGACATATA 965
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93697 CGTTGGCCCCACAATGGCTTGTTGGAAGACTAACAATTCATCTTTGCAGGATTTGAGACATATA 93638

Qy    966 AATTGAACATGAAACAGGGAACCAAGATAACAATTCATCTTCATCCTTGCAGCGATTAACAATTAACATC 1025
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  93637 AAATGAACATGAAACAGGGAACCAAGATAACAATTCATCTTCATCCTTGCAGCGATTAACAATTAACATG 93578
```

```
Qy   1026 CAACCCGGAAGAACCATCACTGTGAAGGACACAGTTGAGGCCTGTGCCAGTGCTGCTGAGAT 1085
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  93577 CAACCCGGAAGAACCATCACTGTGAAGGGTGAGGCCTGTGCCAGTGCTGCTGAGAT 93518

Qy   1086 AGAGATTATGAGAGCTGCGTGAGGCCTTGAAAATGATATGCTGCGTGTTAACACAACA 1145
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  93517 GGAGATTATGAGAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGTTAA--- 93464

Qy   1146 AGCCAATCTGATCCCAGGGTTGAACCTTGGCATCTTTTCAAGGACTGTC 1205
Db  93463 ------- 93464

Qy   1206 CGTGCTATCTCCACCAGAGGGCCGGGAGTCCCCCGCTGCCCCCTACCACCCCTT 1265
Db  93464 ------- 93464

Qy   1266 CACTACCACTCCGGATAC-TTCTCAGCTGTACCCCATCACCAGTTGCCCGTTCC 1324
          ||||||| |||||||||| ||||||||| ||||||||||||||||||||||||
Db  93463 --TACCCACTCCAGATACTTTCTCCAGCTGTACCCCATCACCAGTTGCCCGTTCC 93407

Qy   1325 CGCATCATCATCCTTATCCAGAGCAGAGATTGTGAATCTTCTTCATCCCAACCCAGGCTG 1384
          |||||||||||||||| ||||||||||||||||||||||||||||| |||||||||||||
Db  93406 CGCATCAGCATTCTTATCAAGAGCAGAGATTGTGAATCTTCATCCAACCCAGGCTG 93347

Qy   1385 TGGGCGCCATCATCGGAAGAAGAGGGGCACACATCAAACAGTGGCGAGATTCGCCGGAG 1444
          |||| |||||||||||||||||||||| ||||||||||||||| ||||||||||||||
Db  93346 TGGGTACCATCATCGGAAGAAGAAGGGGCACACATCAAACAGTGGCGAGACTCGCCAGAG 93287

Qy   1445 CCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCATCA 1504
          |||||| |||||||||||||||||||||||||||||| |||||||||||||||||||
Db  93286 CCTCCA------ 93281

Qy   1505 CCCGGCCACCGGAAGCCCAGTTCAAGGCCAGGACGGATCTTTGGGAAACTGAAAGAGAGG 1564
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  93280 ------TCAAGGCCAGGGACGGATCTTCGGGAAACTAAAAGAGAG 93242

Qy   1565 AAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCTT 1624
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |
Db  93241 AAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCCCATATCAGAGTGCTCTCT 93182

Qy   1625 CCACAGCTGCCGGTGATTGGCAAAGTGGCCAAGACCGGCAAGACGCAGAACTAA 1684
          |||||||||||||||||||||| ||||||||||||||||||||||| ||||||||
Db  93181 CCACAGCTGCCGGTGATTGGCAAAGTGGCAAGACTGGCAAGACTGAACTGAACTAA 93122

Qy   1685 CCAGTGCAGAGTCATCGTGCCTGTGACCAAACGCCAGATGAAAATGAAAGGAAGTGATCG 1744
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  93121 CTAGTGCAGAAATCATTGTGCCTGTGACCAAACGCCCGATGAAAATGAAAGGAAGTGATCA 93062

Qy   1745 TCAGAATTATCGGGCACTTCTTTGCTCAGCGAAGGCCGCAAGATCAGGGAAATTG 1804
          |||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db  93061 TCAGAATTATCGGGCACTTCTTTGTTGTCAGCGAAGGCCGCCAGCGCAGTATCAGGGAAACTG 93002

Qy   1805 TACAACAGGTGAAGCGAGCAGCAGGAGAATACCCTCAGGGAGTCGCCTCACAGCGCAGCA 1864
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  93001 TACAAGAGGTGAAGCGAGCAGCAGGAGTAGAAATACCCTCAGGGAGTCGCCTCACAGCGGCAGCA 92942

Qy   1865 AGTGAGG-CTTCCACAGAGGCACCAGCAAAACCAGCCCTTCCAACACCTG 1923
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  92941 AGTGAGGTCTCCCACAGAGGCATCAGCAACATCAGCAACACAGCCCTTCCAACACACCTG 92882

Qy   1924 ACAGAATGAGACAAACGCAGCCAGCCAGGCGGATCGGGAGCAAACCAATCTGAGGAA 1983
          ||||||||| ||||||||||| |||||||||||||||||||||||||||||||
Db  92881 ACAAATGAGACAAACAGCAGCCAGCCAGGCGGAAGCAAACCAAGACCATCTGAGGAA 92822

Qy   1984 TGAGAAGTCTGCGGAGGGGAAGGTTCTGGAAGACTCGCAGAACCCTGAGAGCCGGGGCCGA 2043
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  92821 TGAGAGGTCTGCGGAGGGGAAGGTTCTGGAAGACTCGCAGAATCCAGGGGCCAA 92762

Qy   2044 GGAGGGGCGGGGAAGGTCAGCGAAGGTTGCCAGGATTTGAGCATATA 2103
          |||||  ||||||||||||||||||||||||| |||||||||||
Db  92761 GGAGCG-GGGGAAGGTCAGCGAAGGTTGCCAGATTCCTCTACCTCC 92703

Qy   2104 CAGGGGTTCTGCAGCTTCAGCCATGGCCTTCACCATCACTTCAGCAACCACTTCCTGAACTC 2163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92703 AAATGAACATGAAACAGGGAACCAAGATAACAATTCATCTTCATCCTGGATCTTCCTGAACTC 93637
```

```
Db  91631 ACTTCAGGTGTGTTTTTGCCTCAGCACCTTGCTCTTGTGTTTCTTAGAGATTTGTGAA 91572
          |||||||||||  ||||||||| |||||||||||||||||||||||||||||||||
Qy   3233 GCTGATAGTTGGAGCATTTTTTATTTTTTAATTAATGAGTTGGAAAAAATAAGA 3292
          |||||||||||||||    |||| |||||||||||||||||||||
Db  91571 ACTGATAGTTGGAGCA----TTTATTTCTTAATAAATGAGTTGGAAAAAATAAGA 91516

Qy   3293 TATCAACTGCCAGCTGGAGAAGGTGACAGTCCAAGTGCAACAGCTGTTCTGAATTGT 3352
          |||||| ||| ||||||||||||||| |||||| |||||||| ||||||||||||||
Db  91515 TATCAACAACAGCCTGGAGAAGGTGAGAGTCCAATGCTGCAACAACTGTTCTGAATTGT 91456

Qy   3353 CTTCCGCTAGCCAAGAACCNATATGGCCTTCTTTGGACAAAATGTTATT 3411
          |||| ||||||||||||||| |||||||||||||    |||||||||||||
Db  91455 CTTCTGCTAGCCAAGAACCTATATGGCCTTCTTTTAGACAACAACCTTGAAAACATTATT 91397
```

RESULT 14
BC023758
LOCUS   BC023758   3557 bp   mRNA   linear   ROD 16-APR-2003
DEFINITION Mus musculus RIKEN cDNA C330012H03 gene, mRNA (cDNA clone
            IMAGE:5354659), partial cds.
ACCESSION   BC023758
VERSION   BC023758.1 GI:23958572
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3557)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE   22388257
PUBMED   12477932
REFERENCE   2 (bases 1 to 3557)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (05-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S.J., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
```

```
Db  92702 CAGGACTTCTGCA-GCTTCAGCAATCCACTTCACCACCACTCAGATCTCTCCTTAACTC 92644
          |||||||||||  ||| ||||||||||||||||||||||||||||||||||
Qy   2164 CCACGACGCTATCCCTTTAGTTGACTAACAATAGGTGAACGTGTTCAAAGCCAAGCAAA 2223
          |||| || |||| ||||| ||||| ||||||||||||||||||||||||||| |||||
Db  92643 CCACAATGCTGTACCTTTTGGTTGAAGTAACATAGGTGAACGTGTTCAAAGCCATGCAA 92584

Qy   2224 ATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATTAGAAAGGG 2283
          ||||||||| |||||| ||||||||    |||||||||||||||| |||||||| |||||
Db  92583 ATGCACACCCCTTTTTTGTGTGGCAAA----TCTCTGTACATGTGTGTGCACATTAGAAGGGG 92528

Qy   2284 AAGAGTTAAGATATGTGGCCTGTGGTTACACAGGGTGCCTGCAGCGGTAATATATTT 2343
          | || ||||||| |||||||||| |||| |||||||||||||||||||||| ||||||
Db  92527 ACGATGTAAGATACGTGGCCTGCGGGTTACGCAGGGTGCCTGCAGCGGTAATATGTTTT 92468

Qy   2344 AGAAATAATATATCAAATAACTAACTAACTAATTCCAATTTTTAATCAATTATTAATT 2403
          ||||||| ||||||||| ||| |||||||||||||||||||||||||| || ||||||
Db  92467 AGAAATAACATATCAAATCAATAACTAACTAATTCCAATTTTTAATCAATATTAATTTTTT 92408

Qy   2404 TTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTGGGGAGTC 2463
          ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db  92407 TTCTTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGTCT----TTCTTGGGGAGT 92352

Qy   2464 TCACGGTGTAGAGGAGAGGGAGTTTGAGGCCACCCGCACACAAAATTCACCCGAGGGAAATCTC 2523
          ||||||||||||||||||||||||||||| | || | |||||| ||||||||||||||||||
Db  92351 TCACGGTGTAGAGGAGAGGGAGTTTGAGGTGACCTACACAAATTCACCCGAGGGAAATCTT 92292

Qy   2524 GTCGGAAGGACACTCACGGCAGTTCTGGAATCACCTGTATGTCAAACAGAAGGGATACCG 2583
          |||||||||||||||| |||||||||| ||| ||| |||||| || |||||||||||| |
Db  92291 GTCGGAAGGACACTCACAGCAGTTCTGTATATCGTGTATATCAACGAGGAGGATACCA 92232

Qy   2584 TCTCCTTGAAGAGAAACTCTGTCACTCCTCAGCCTGTCTAGCTCATCACACCCATTTCT 2643
          |||||||||| || ||||||||||||||||| ||||||||||| |||| ||||||||
Db  92231 TCTCCTTGAAAAGGAAACTCTGTCACTCCTCATCGCCTGTCTAGCTTATCATCCATTTCT 92172

Qy   2644 CTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTATATAATCTC--TGTCTC 2701
          |||||||||| || |||||||||||||||| ||||||| ||||| | |||| ||||||
Db  92171 CTTTGCTTCATGGGGTTTTAAACTGGTTTTTCGCATACTGCTATATTCTCTATCTCTC 92112

Qy   2702 TCTCTGTTTATCTCCCCCTCCCCTCCCCTTCTTCTCCATCTCCATTCTTTTTGAAT 2761
          |||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db  92111 TCTCTGTTTATCTCCCCCTCCCCTCCCCTACTTCTCCTCCATCTCCATTCTTTTTGAAT 92052

Qy   2762 TTCCTCATCCTCCTCCATCTCAATCCCGTATCTAGCCA------CCCCCCCCCCCAGG 2813
          | ||||||||||||||||| |||| | |||||| ||      |||||| |||||||
Db  92051 TCCCTCATCCTCCTCCATCCATATCTACGCAACACCACCCCTGCCCCCGCCCCCAAG 91992

Qy   2814 CAAAGCAGTGCTCTGAGTATCACATCACACAACAAAGGAACAAAAGCCGAAACACACAAACCA 2873
          ||||||||| |||||||||| ||| || ||||||| |||||||||| |||||||||| |||
Db  91991 GAAAGCAGTGCTTTGAGTATCCCACCAACAACAGAACAAAAGCAAAACAACACACGTCA 91932

Qy   2874 GCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTT 2933
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||
Db  91931 GCCTCAACTTACGCTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCATT 91872

Qy   2934 TTGGAAGAGGAAAACAGAGAACCACCAAACCACCAAACAACAAAGAAAAATTC 2993
          |||||||| ||||||||||||||||| |||||| |||||| |||||||||||||
Db  91871 TTGGAAGAGAAAACAGAGAACCACCAAACCAACAATCAACCAAACAAGAAAAAATTC 91812

Qy   2994 CACAATGAAAGAAGAATGTATTTTGTCTTTTGTCATTTGGTGTATAAGCCATCAATATTCAG 3053
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  91811 CACAATGAAAGAAGAATGTATTTTTGTCTTTTGTCATTTGGTGTATAAGCCATCAATATTCAG 91752

Qy   3054 CAAAATGATTCCTTCTTTAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTT 3113
          |||| | ||| ||| ||| ||||||||||| | ||||||||| ||||| ||||||
Db  91751 CAAATTATTTCTTTCTTTCTTTTTTAAAAAAATGTGAAGGAAAGTAGCAATTTATACAAGGTT 91692

Qy   3114 GTTGGCCAGGGCGTTAAATTCACAGATTTTTTTACGAGAAAAACACAGAAGAGAGCT 3173
          ||||| ||||||| ||||||| || ||||||| || | ||| |||||| |||| |||
Db  91691 GTTGGGCCAGGGCGGTTAAATTTACCGATTTTTATTATCACCACAGAAAACACAGAAAAAGCT 91632

Qy   3174 ACCTCAGGTG-TTTTTACCTGCACCTTGCTCTTGTGTTTCCCTTGTTGCTTAGAGATTTTGTAAA 3232
          |||||||||| |||||| || |||||||||||||||||||||  ||  |||||||||| ||||
```

Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 55 Row: o Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES   Location/Qualifiers
     source   1..3557
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5354659"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match   63.5%;   Score 2165;   DB 10;   Length 3557;
Best Local Similarity   83.0%;   Pred. No. 0;
Matches 2820;   Conservative   0;   Mismatches   476;   Indels 100;   Gaps   27;

Qy   58 GGGAAGAGACGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCCCGTCACCGCC 117
         |||||||||| ||||||||||||||||| ||| |||||||||||||||||||||||||||
Db   23 GGGAAGAGACGGGATGATGAACAAGCTGTACATTGGGAACCTGAGCCCCCCGTCACCGCC 82

Qy   118 GACGACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGGACAGGTCCTGCTG 177
         |||||||||||||||||||||||||| ||||||||||| ||||||||||||||||| ||||
Db   83 GACGACCTCCGGCAGCTCTTCGGGGACAGGAAGCTGCCCCTGGCGGGGACAGGTCCTACTC 142

Qy   178 AAGTCCGGCTACGCCGCCTTCGTTGGACTACCCCGACCCAGAACTCGGTGGACACGTCAACA 237
         |||||||||||||||||||||||||||||||||||||||||||| ||||||||||| |||||
Db   143 AAGTCCGGCTACGCCGCCTTCGTTGGACTACCCCGACCCAGAACTCGGGCCATCGCGCCACC 202

Qy   238 ACCCTTCTCGGGTAAAGTGCAAGTTGCATGGGAAAATCATGGAAGTTGATTACTCAGTGTCT 297
         |||||| |||||||||||||||| ||| || |||||||||||||||||||||| |||||||
Db   203 ACCCTTCTCGGGTAAAGTGCAAGTTGCATGGGAAAAATCATGGAAGTTGATTACTCAGTGTCT 262

Qy   298 AAAAAGCTAAGGAGCAGGAGAGAAATTCAGATTCGGGACAAACATCCCTCCTCACCTGCGGAG 357
         |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   263 AAAAAGCTAAGGAGCAGGAGAGAAATTCAGATTCGGGACAAACATCCCTCCTCACCTGCGGAG 322

Qy   358 GTGTTGGATGAGACTTTTGGCTCAATATGGGACAAGTGGAGAAGTGGAACAAGTCAACACA 417
         ||||||||||| ||||||||||||| |||||||||| |||||||||||||| |||||||||
Db   323 GTGTTGGATGGGCTGTTGGCTCAGTGAATATGGGACAAGTGGAGAAGTGGAACGTGTCAACACA 382

Qy   418 GACACAGAGAACCGCCCGTTGTCACGTTCAACGTCACATATGACAACAGAGAAGGCAAAATAGCC 477
         ||||||||||||||||||| ||||||| ||||||| |||||||||||||||||||||||||||||
Db   383 GATACAGAGAACTGCCCGTTGTGGCTCAACGTCACATATGACAACAGAGAAGGCAAAGCTAGCT 442

Qy   478 ATGGAGAAGTCAAGCGGGGCATCATCCGGGCATCGGCCTTTGAGAACTACTACCATCCCG 537
         ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443 ATTGAGAAGTCCAGTGGGGCTCAGTTTGATGGGCATCGGCCTTTGAGAACTACTACCATCCCG 502

Qy   538 GATGAAGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCGAGCCCAGCGTGGGGACCACTCTTCC 597
         ||||||||||||||||||||||||||||||||||||||||||||||||||||----------
Db   503 GATGAAGAGGTGAGCTCCCCTTCGCCCCCTTCGCCCGAGCCCAGCGTGGGG----TGCC 541

Qy   598 CGGGAGCAAGGCCACGCCCCTGGGGGCATCAGGCCACAGATTCAGGCCACAGATTCCCGCTG 657
         ||||||||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db   542 CGGGAACAAGGCCACGCCCCGGGGAGCTCTTCTCAGGCCACAGATTCAGGCCACAGATTCCCGCTG 601

Qy   658 CGGATCCTGGTCTCCCCACCCAGTTTGTTGGTGCCATCATCGGGCATTTTTTGCTAGCCAGGATTCAGG 717
         ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   602 CGGATCCTGGTCTCCCCACCCAGTTTGTTGGTGCCATCATCGGGCATTTTGTTGCTAGCCAGGATTCAGG 661

Qy   718 AAGAACATCACTAAGCAGACGACCCAGTCCCGGTGGTGTAGTATCCATAGAAAGAGAACTCTGGA 777
         |||||||||||||||||||||||||||||||| |||||| |||| ||||||||| |||||||||
Db   662 AAGAACATCACTAAGCAGACGACCCAGTCCCGGGTAGACATCCCAGAGAAAGGAGAACTCTGGG 721

Qy   778 GCTGCAGAGAAGCCTGTCACCATCCATGCATCCTCACCCAGAGGGGACTTCGAAGCATGCCGGC 837
         |||||||||||||||||||||| ||||||| |||||||||||||||||||||||||||||||||
Db   722 GCTGCAGAGAAGCCTGTCACAATCCATGCATCCTACCCCAGAGAAGGGACATCAGCAGCATGCCGC 781

Qy   838 ATGATTCTTGAAATCATGCAGAAAGAGCAGATGAGACGAAACTGGCCGAAGAGATTCCT 897
         |||||||| |||| ||||| |||||||||| ||||||||| |||||| ||||||||||||
Db   782 ATGATTCTTGAGATTATGCAAAAAGAAGAGCTGATGAGACCAAACTGGCCGTGAGAGGTTCCT 841

Qy   898 CTGAAAATCTTGCACACAATGGCTTGTTGAAGACTGATTGGAAAAGAGAGGCAGAAAT 957
         |||||| |||||| ||| |||||||||||||||||| |||||||||||| |||||||
Db   842 CTGAAAATCCTGCGCCCACAATGGCTTCGTTGGAAGACTGATTGGCAAAAGAAGGCAGAAAC 901

Qy   958 TTGAAGAAATTGAACATGAAACAGAGGACCAAGATAACAATCTCATCTTTGCAGGATTTG 1017
         |||||||| |||||||| ||||||||||| ||||||||| ||| ||||||||||||||||
Db   902 CTGAAGAAATAGAACATGAAACAGAGGACCAAGATATACCAATCTCATCCTTGCAGGATTTG 961

Qy 1018 AGCATATACAACCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGGCTGTGCCAGT 1077
         |||||||| |||||||||||||||||| |||||||||||||||||||||| |||||||||
Db   962 AGCATTTATAACCCCGAGAGAACCATCACCCGTGAGGGGGCACCCATTGAGGCCTGTGCCAAT 1021

Qy 1078 GCTGAGATAGAGATTATGAAGAAGCTCGCTGAGGGCCTTGAAAATGATAATGCTGGCGTGTT 1137
         ||||||||||||||||||||||||||| ||||||||||||||||| ||||| ||||||||||
Db 1022 GCTGAGATAGAGATTATGAAGAAGCTCCGAGAGAGCCTTTGAGAACGACATGCTGGCGCATCTT 1081

Qy 1138 AACCAACAAGCCAATCTGATCCCGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACA 1197
         ||||||||| ||||||||||||||||||||| ||||||| |||||||||||| |||||
Db 1082 AACCAACAAGCCAATCTGATCCTGATCCTCACCCCAGTGCCTAAACCTCAGCACTTGGCATCTCTTTTCGACT 1141

Qy 1198 GGACTGTCCGTGTATCTCCACCAGCAGAAGCTGCCCCTGGCGGGACAGGTCCTGCTG 1257
         |||||||||||| |||||||||||||||||||||| ||||||||||||||||| |||
Db 1142 GGACTGTCTGTGCTTCCTCCACCAGCAGAAGCTGCCCCGTGGAGTTCCCCCCAGTCCTCCTAT 1201

Qy 1258 CACCCCTTCACTACCTACCACTCCGATACTTCTCCAGCGCCTCGTCCATCACCAGTTTGGC 1317
         ||||||||| ||||| ||||||||||||||||| |||||||||||||||||||||||||||
Db 1202 CACCCCTTTGCTACCACCACTCCGATACTTCTCCAGTCTGTCCATCACCATTTCGGC 1261

Qy 1318 CCGTTCCCGCATCATCACTCTTATCCGAGACAGGAGAGATTGTGAATCTTCATCCCACC 1377
         |||||||| ||||||||||| |||||||||||||||||||||| |||| ||||||||||
Db 1262 CCATTCCCACCACATCATCATCACTCACTCCTACCCAGGAGCAGACTGTAAGTCTCTTCATCCCAACC 1321

Qy 1378 CAGGCTGTGCGGCGCCATCATCGGGCGGAAGAGGGGGCACACTCAAACAGCTGGCGAGATTC 1437
         |||||||||| ||||| |||||||||||||||||||| |||||||||||| |||||||||||
Db 1322 CAGGCTGTGTGGTGCTATCATCGGGCGTATCATCGGGAGAAGGGGGCACACACAAACAGCTCGCTGATTT 1381

Qy 1438 GCCGGAGCCCTATCAAGATTGCCCCTGCGGAAGGTCAGCGGAAAGGATGGTC 1497
         |||||||||||||||||||| ||||| ||| || |||||| ||||||||| |
Db 1382 GCTCGTGCCTTCCATCAAGATTGCTCCAGCATGTCCAGTGTCCAGAGGATGGTC 1441

Qy 1498 ATCATCACCCGGCCACCCGGAAGCCCAGGCCCAGGCACCGGAAACTG 1557
         |||||||||| |||||||||||||| ||||||||||| |||||||||
Db 1442 ATCATCACTGGTCCTCCTGAAGCCCCGGCCCAGTTTAAGGCTGGAGGCGAGTTGGGAAACTG 1501

Qy 1558 AAAGAGGAAAACTTCTTCAAGATTCCTGAGAACTACTACCCG 1617
         |||||||||||| |||||| |||||| |||||||||||||||
Db 1502 AAGGAGAAAAACTTCTTTAATCCAAAGAAGAGAAGAGAAGTGAGCTGAGCGGCCCACATCCGAGTC 1561

Qy 1618 CCCTCTTCCACAGCTGGCTCGGGTGATTGGGAAAGGTGGCAAGA 1677
         ||||||| |||||||| ||||||||||| |||||||||||||||
Db 1562 CCCTCGTCGACCGGCTGGCCGGGTGATTGGGAAAACCGTGCAAGGAA 1621

Qy 1678 AACTTAACAGTGCAGAAGTGCATCGTGCCTCGTGACCAAACGCCAGATGAAATGAGGAA 1737
         ||||||| |||||||| |||||||| |||||||||| ||| |||||| |||||||| ||
Db 1622 AACTTGACAAGTGCAGAAGTGCAGAAGTTATCGTGCCTCGTGACCAAACGCCAGAGAATGAGGAA 1681

Qy 1738 GTGATCGTCAGAATTATCGGGCCACTCTTCTTCGTCCTAGCCAAGATCAGG 1797
         ||||||| |||||| ||||||||||| ||||||||||||| |||||||||||
Db 1682 GTGATCGTCAGAATTATCGGGCCATTTTTTGCTAGCCAGGGTCAACCGCAAGATCAGG 1741

Qy 1798 GAAATTGTACACAGTGAAGCAGCAGAAATACCCTCAGGGAGTCGCCTCACAG 1857

```
Db  2798  CACACAAGGAACAAATGGAACACACAGACCAGCCTCAGCTTACACTCGGTTACTCAAG 2857
          |||||||||| |||| ||||| |  |    | ||||||||  |||  ||||| |||||
Qy  2901  AGAACAAGAGTCAATGGTACTTGTCCTACCGTTTGGAAGAGGAAAACAGGAACCCACCA 2960
Db  2858  CGAACAAGAGTCGGCGGTACTTGTCCCAGC-ATTCGGAAGAGGAGAAGCAGGATCCCAGCA 2916

Qy  2961  AACCAACCAATCACCAACAAGAAAAATTCCACAATGAAAGAATGTATTTTGTCTT 3020
          |||  |  |  | |||||||  | |||||| |||| |||| || |    |||| |
Db  2917  AATAAACCAGCCAACCAAACAAG---AAAATCCCACAAAGAAGAATGGA---CTCTCT 2971

Qy  3021  TTGCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATGATTCCTTTCTTTAAAAAAA 3080
          |||||| |||||  ||| ||||||||  || ||| |||| || |  | ||||||||| ||
Db  2972  CTCCATCTTGGTATACAGCCATCAACATTTAGTGAAATTCTTCTTTCTTTAAAAATTT 3031

Qy  3081  AAATGTGGAGGAAAGTAGAAAATTTACCAAGGTGTTGTTGCCCAGGGCGTTAAATTCACAGA 3140
          || ||||||||||| || |||| || || ||||||||||  |||||||||||||| ||||||
Db  3032  AAAGTGGAGGAAAATAGCAATTTATACGAGGTTGTTGGCCCAGGGCGTTAAATTTACAGA 3091

Qy  3141  TTTTTTTAACGAGAAAAACACACACAGA--AGAAGCTACCTCAGTG-TTTTTTACCTCAGC 3196
          ||||||||| ||||||||||||||| ||  | |||||||||||| | |||||||||||||||
Db  3092  TTTTTTTAAATGAGAAAAACACACAAATAAAAAGCTACCTCAGGTGTTTTTTTTACCTCAGC 3151

Qy  3197  ACCTTGCTTCCCTTGTTCCCTTAGAGATTTTGTAAAGCTGATAGTAGTGGAGCATTTTTTA 3256
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||| |||||
Db  3152  ACCTTGCTTCCCTTGTTCCCTTAGAGATTTTGTAAAACTGATAGTAGTGGAGCA-TTTTTTA 3210

Qy  3257  TTTTTTTAATAAAAATGAGTTGGAAAAAAAAAGAGATATCAACTGCCAGCGTGGGAAGAGG 3316
          ||||||||||||||||||||||||||||||| |||  |  ||||||||||| |||  |||||
Db  3211  TTTTTTTAATAAAAATGAGTTGGAAAAAAAAAAAAAG--ATGAGCTGCCAGCCCCAAGCAGG 3268

Qy  3317  TGACAGTGCCAGTGTGCAAAGCTGTTCTGAATTGTCTTCCGCTAGCCAAGAACNATAT 3376
          ||  |   ||| ||||||  ||||||||    ||||||||||||||||||| | |||||
Db  3269  CGGCAGCCCAGGTGTGCCGCAGCTCTTC-----TGTCTTCCGCTAGCCAAGAGCCTACGT 3323

Qy  3377  GGGCCTTCTTTTGGCAAACCTTGAAAAATGTTTATT 3412
          || ||||||| |||| |||||| ||||||||||||||
Db  3324  GGCCTTCTTGTGTGGACAAACCTAAAAATGTTTATTTT 3359
```

```
Db  1742  GAAATTGTACAGCAGGTGAAGCAGTGAAGCAGCAGCAGAGAGATACCCTCAGGGAGTGCCCCACAG 1801
Qy  1858  CGCAGCAAGTGAGGCTCCCACAGGGCACCAGCAAAACAACGATGAATGTAGCCCTTCAA 1917
Db  1802  CGCAGCAAGTGAGGCTCCCACAG--CACCAGCAAGCAACCGATGAATGTAGCCCTCCCAA 1859
Qy  1918  CACCTGCTCAGAATGAGACCAAAACGCAGCCAGCAGACGGAGCAAACCAAAGACCAT-C 1976
Db  1860  CACCTGCACAG-ATGAGACCAAAC---AGCCAGCAGATCGGAGCAAGCAAACCAAAGAGCATCC 1915
Qy  1977  TGAGGAATGAGAAGTCTGCGAGGCGCGGCGCCTGCGAGGGCCCTGAGACCCCCAG 2036
Db  1916  CGAGGAGTGCCAGTCTGCAGAGCGCCATCCCTGC----AGACCTCTACACATCCTG 1970
Qy  2037  GGGCCGAGGAGGCGGGGGGAAGGTCAGCCAGGTTGCCAGAACCACCG--AGCCCGCCT 2094
Db  1971  GGATCCAGGAGGGGCACAGGGAAG--GCCAGGTGCTGCCAGAAACACCGCTTGGCCTGCCC 2027
Qy  2095  CCGCGCCCCCAGGGCTTCTGCGAGGCTT--CAGCCATCCACTTCCACCATCTCGGATCT 2152
Db  2028  CCAGCTTCCCCTGGCTTCTGCGAGGCGATAACAGCCATCCACTGCCAATCCAGAT-T 2086
Qy  2153  CTCCTGAACTCCCACGAGCGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAA 2212
Db  2087  CTCCTCAGTTCCCAGACGCGCTATCCCTTTCGGTTGAACTAACATAGGTGAAACATGCTCAA 2146
Qy  2213  AGCCAAGCAAAAATGCACACACCCTTTTC----TGTGGCAAATCGTCTGTACATGTGTGT 2268
Db  2147  AGCAAAGCAAAATTCCTAGCCGTTTCTTTGTTGTGGAAAGTTGTCTGTACATGTATGT 2206
Qy  2269  ACATATTAGAAAGGAA-GATGTTAAGATATGTGCCCTGGGGTTACACAGAGGTGCCTGC 2327
Db  2207  ACATATCAGAAGGAAGGATGTTAAGAAATGTGCCCTGGGGTTACACAGGGTGCCTGC 2266
Qy  2328  AGCCGTAATATATTTTAGAAATATATATCAACTAACTCCAATTTTAAATC 2387
Db  2267  AGCGTTAATATATTTTAGAAATAATAATCAAATAACTCCACTAACAACAATTTTACTC 2326
Qy  2388  AATTATTAATTTTTTTTCTTTTAAAGAAAGCAAGCTTTTCTAGAACTTTAAAGAATA 2447
Db  2386  AGTGATTAATTTTGTTTCTTTTTGTTTCTTTTCTAGAACTTTCTAGAACTTCCAAGAATA 2386
Qy  2448  AAGTCTTTGG-GAGGTCTCACGGTGTAGACAGAGCTTTGAGGC-CACCCGCACAAAATT 2505
Db  2387  AAGTATTGGGTGAGGTCCACTGGTATAGAAAGGAACTTTGAGCCACCCACCACAAAATT 2446
Qy  2506  CACCCAGAGACGGAAATCTCGTCGGAAGGACACTCACGGACAGTTCTGGATCACCTGTGTATG 2565
Db  2447  CACTTGGAGGACGGAAATCTTGTC-AAGGACACTTAATGCAGTTGTGGGTCACCTGTGTCTG 2505
Qy  2566  TCAACAGAAGGGATACCGTCTCCTTGAAGAGAGAAACTCTGTCACTCCTCCATGCCTGTCTA 2625
Db  2506  TCAACAGAAGGGATACCG--TCCCTGACGAGGGGGGCCCTAGGCCCTGGTCACCTGTTCTC 2563
Qy  2626  GCTCATACACCCATTCTCCT---------------------TGCTTCACAGGTTTTAA 2663
Db  2564  CCACACACCCATTGTCCTGTGTCTCCAGTTCACAGAGTTTTTAAACTGATGATTTTGGGTTTTTGT 2623
Qy  2664  ACTGGTTTTTTGCATACATGCTATATAATATTCTCGTCTCTCGTTCGTTATCTCCTCC 2723
Db  2624  TTTTGTTTTTTGCATACTGCTATATATGTTCTGTCTCTCTC------CTTATCTCCCTCCT 2677
Qy  2724  CTCCCCTCCCCTTCCTTCTTCTCCATCTCTTTTGAATTCCTCATCCCTCCATCTCAAT 2783
Db  2678  AACCTTCCTCTGTGTCTTCTTTACCCAAAATTCTTTGAGTCCTTGGTCCTTGTTCTTTCTGCTC 2737
Qy  2784  CCCGTATCTACGGCACC---CCCCCCCCCAGGCAAAGCTGCTCTGAGTATCACATCA 2840
Db  2738  CCCGTATCTCATGCACCCCCCTCCCCCTCCCCCCAGGCAAAGCCTCCGAATATCATACCA 2797
Qy  2841  CACAAAAGGAACAAAAGCGAAACACACAAACCAGCCTCAACTTACACTTGGTTACTCAAA 2900
```

```
                    /gene="IMP-2"
                    /note="alternative start-codon gives an additional
                    N-terminal methionine"
CDS                 50..1846
                    /gene="IMP-2"
                    /note="binds multiple sites in IGF-II mRNA 5'UTR"
                    /codon_start=1
                    /product="IGF-II mRNA-binding protein 2"
                    /protein_id="AAD09827.1"
                    /db_xref="GI:4191610"
                    /translation="MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDY
                    PDQNWAIRAIETLSGKVELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLA
                    QYGTVVENVEQVNTDTETAVVNVYATREAKIAMEKLSGHQFENYSFKISYIPDEEVS
                    SPSPPQRAQRGDHSSREQGHAPGGTSQARQIDFPLRIIVPTQPVGAIIGKEGLTIKNI
                    TKQTQSRVDIHRKENSGAAEKPVTIHATPEGTSEACRMLEIMQKEADETKLAEEIPL
                    KILAHNGLVGRLIGKEGRNLKKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACA
                    SAEIEIMKKLREAFENDMLAVNQQANLIPGLNLSALGIPSTGLSVLSPPAGPRGAPPA
                    APYHPFTTHSGYFSSLYFPHQFGPFPHHHSYPPQEIVNLFIPTQAVGAIIGKKGAHIK
                    QLARPAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQRIFGKLKEENFPNPKEEVK
                    LEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEYVIVPRDQTPDENEEVVRIIGHFFA
                    SQTAQRKIREIVQQVKQQEQKYPQGVASQRSK"

ORIGIN

Query Match      58.0%;    Score 1979.8;    DB 9;    Length 2010;
Best Local Similarity   99.4%;    Pred. No. 0;
Matches 1987;   Conservative   0;   Mismatches   12;   Indels   0;   Gaps   0;

Qy    35  GCCGGGGGAGCGGCGGGGCTCTGGGGAAGAGACGGATGAGAACAAGCTTTACATCGGGA  94
          |||||| ||||| ||| ||||| |||||||||||||||||||||||||||||||||||
Db    12  GCCGTAGGGAGCCGGCCGGGCCTCAGGGAAGAGACGGATGAGAACAAGCTTTACATCGGGA  71

Qy    95  ACCTGAGCCCCGCCGTCACCGCCGACCTCCGGCTACCGGCTACCCCGACCAGCTGC  154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72  ACCTGAGCCCCGCCGTCACCGCCGACCTCCGGCTACCGGCTACCCCGACCAGCTGC  131

Qy   155  CCCTGGGGGGACCAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGA  214
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132  CCCTGGGGCGGACCAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCCGACCAGA  191

Qy   215  ACTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAAGTGGAATTGCATGGGAAAATCA  274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192  ACTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAAGTGGAATTGCATGGGAAAATCA  251

Qy   275  TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGAAAATTCAGATTCGAAACA  334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   252  TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGAAAATTCAGATTCGAAACA  311

Qy   335  TCCCTCCTCACCTGCAGTGGAGAGTGTTGGATGGACTTTGGCTCAATGGGACAGTGG  394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312  TCCCTCCTCACCTGCAGTGGAGAGTGTTGGATGGACTTTGGCTCAATGGGACAGTGG  371

Qy   395  AGAATGTGAACAAGTCAACACAGAACACAGAAACCGCCGTTGTCAACGTCACATATGCAA  454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372  AGAATGTGAACAAGTCAACACAGAACACAGAAACCGCCGTTGTCAACGTCACATATGCAA  431

Qy   455  CAAGAGAAGAAGCAAAAATAGCCATGGAGAAGCTAAGCGGGCCATCAGGCGGGCACTACT  514
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   432  CAAGAGAAGAAGCAAAAATAGCCATGGAGAAGCTAAGCGGGCCATCAGGCGGGCACTACT  491

Qy   515  CCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGCTGAGCCTTCGCCCCCTTCAGCGAG  574
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   492  CCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGCTGAGCCTTCGCCCCCTTCAGCGAG  551

Qy   575  CCCAGCGTGGGACCACTCTTCCCGCTGGGGGCAAGGCCACGCGGGGGCACTTCTCAGG  634
          ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   552  CCCAGCGTGGGGACCACTCTTCCCGCTGGGGGCAAGGCCACGCGGGGGCACTTCTCAGG  611

Qy   635  CCAGACAGATTGATTTCCCGCTGCGGATCCTGCGTTGGTCCCCAGTTGTTGGTGCCATCA  694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   612  CCAGACAGATTGATTTCCCGCTGCGGATCCTGCGTTGGTCCCCAGTTGTTGGTGCCATCA  671

Qy   695  TCGGAAAGGAGGGGCTTGACCATAAGACAACATCACTAAGCAGACACCAGTCCCGGGTAGATA  754
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672  TCGGAAAGGAGGGGCTTGACCATAAGACAACATCTAAGCAGACACCAGTCCGGGTAGATA  731

Qy   755  TCCATAGAAAAGAGAACTCGGAGCTGCAGAGAAGAGCCTGTCACCATCATCGCACCCCAG  814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   732  TCCATAGAAAAGAGAACTCGGAGCTGCAGAGAAGAGCCTGTCACCATCATCGCACCCCAG  791

Qy   815  AGGGGACTTCTGAAGCATGCGCATGATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGA  874
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   792  AGGGGACTTCTGAAGCATGCGCATGATGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGA  851

Qy   875  CCAAACTAGCCGAGAGATTCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC  934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   852  CCAAACTAGCCGAGAGATTCCTCTGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC  911

Qy   935  TGATTGGAAAAGAAGGCAGAAATTGAACATGAACATGAACATGAAAACAGGGACCAAGATAA  994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   912  TGATTGGAAAAGAAGGCAGAAATTGAACATGAACATGAACATGAAAACAGGGACCAAGATAA  971

Qy   995  CAATCTCATCTTGCAGGATTTGAGCATATACAACCCGAAAGAACCATCACTGTGAAGGCCT  1054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972  CAATCTCATCTTGCAGGATTTGAGCATATACAACCCGAAAGAACCATCACTGTGAAGGCCT  1031

Qy  1055  GCACAGTTGAGCGCCTGTGCCAGTGCTGAGATAGTAGAAGCTGCGTGAGGCCT  1114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1032  GCACAGTTGAGCGCCTGTGCCAGTGCTGAGATAGTAGAAGCTGCGTGAGGCCT  1091

Qy  1115  TTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATTGATCCCAGGGTTGAACCTCA  1174
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  1092  TTGAAAATGATATGCTGGCTCGTTAACCAACAAGCCAATTGATCCCAGGGTTGAACCTCA  1151

Qy  1175  GCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGAAGCTGC  1234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1152  GCGCACTTGGCATCTTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGAAGCTGC  1211

Qy  1235  GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTCGGACTACTTCTCCAGCC  1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1212  GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTCGGACTACTTCTCCAGCC  1271

Qy  1295  TGTACCCCATCACCAGTTGGCCCAACCCAGGCTGTGGGCCGCATCATCGGGAAGAGGGGCAC  1354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1272  TGTACCCCATCACCAGTTGGCCCAACCCAGGCTGTGGGCCGCATCATCGGGAAGAGGGGCAC  1331

Qy  1355  TTGTGAATCTCTTCATCCCACCCAGCCTTCGCCGGAGATTCGCCCCTGCGGAAGGCC  1414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1332  TTGTGAATCTCTTCATCCCACCCAGCCTTCGCCGGAGATTCGCCCCTGCGGAAGGCC  1391

Qy  1415  ACATCAAACAGCTGGCGAGATTGCCCCGGAAGCCCACGGCCAGTTCAAGGCCC  1474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1392  ACATCAAACAGCTGGCGAGATTGCCCCGGAAGCCCACGGCCAGTTCAAGGCCC  1451

Qy  1475  CAGAGACGTCAGCGGAACAAGGATGGTCATCATCACCGGGCCACCCGGGAAGCCCAGTCAAGGGCCC  1534
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1452  CAGAGACGTCAGCGGAACAAGGATGGTCATCATCACCGGGCCACCCGGGAAGCCCAGTCAAGGGCCC  1511

Qy  1535  AGGGACCGATCTTTGGGAAACTGAAAGAGAAAACTTCTTTAACCCCAAAGAAGAAGTGA  1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1512  AGGGACCGATCTTTGGGAAACTGAAAGAGAAAACTTCTTTAACCCCAAAGAAGAAGTGA  1571

Qy  1595  AGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGCCGGTGATTGGCAAAGGTG  1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1572  AGCTGGAAGCGCATATCAGAGTGCCCTCTTCCACAGCTGCCGGTGATTGGCAAAGGTG  1631

Qy  1655  GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAGTACGTCATCGTGCCTCGTGACC  1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1632  GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAGTACGTCATCGTGCCTCGTGACC  1691

Qy  1715  AAACGCCAGATGAAATGAAGAAGTGATCGTCAGAATTATGGGACTTCTTTGCTAGCC  1774
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1692  AAACGCCAGATGAAATGAAGAAGTGATCGTCAGAATTATGGGACTTCTTTGCTAGCC  1751

Qy  1775  AGACTGCACAGCGCAAGATCAGAGAAATTGTACAACAGGTGAAGCAGGAGCAGGAGAAAT  1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1752  AGACTGCACAGCGCAAGATCAGAGAAATTGTACAACAGGTGAAGCAGGAGCAGGAGAAAT  1811
```

```
Qy   1835  ACCCTCAGGGAGTCGCCTCACAGCGCGAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACA  1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812  ACCCTCAGGGAGTCGCCTCACAGCGCGAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACA  1871

Qy   1895  ACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGAT  1954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872  ACGGATGAATGTAGCCCTTCCAACACCTGACAGAATGAGACCAAACGCAGCCAGCCAGAT  1931

Qy   1955  CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCCAGGGACTCT  2014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932  CGGGAGCAAACCAAAGACCATCTGAGGAATGAGAAGTCTGCGGAGGCGGCCCAGGGACTCT  1991

Qy   2015  GCCGAGGCCCTGAGAACCC  2033
           |||||||||||||||||||
Db   1992  GCCGAGGCCCTGAGAACCC  2010
```

Search completed: July 21, 2004, 03:45:54
Job time : 13026 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      July 20, 2004, 20:33:42 ; Search time 1237 Seconds
                                        (without alignments)
                                        1717.746 Million cell updates/sec

Title:            US-09-270-437D-6
Perfect score:    3412
Sequence:         1 ggcagcggaggaggcgagga........aacctgaaatgttattt 3412

Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:         3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_29Jan04:.*
           1: geneseqn1980s:*
           2: geneseqn1990s:*
           3: geneseqn2000s:*
           4: geneseqn2001as:*
           5: geneseqn2001bs:*
           6: geneseqn2002s:*
           7: geneseqn2003as:*
           8: geneseqn2003bs:*
           9: geneseqn2003cs:*
          10: geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	DB ID	ID	Description
1	3411	100.0	3412	3	Aaz36152	AAZ36152	DNA encod
2	3320.6	97.3	3694	4	Aas26150	AAS26150	Human cDN
3	3320.6	97.3	3694	3	Abx73491	ABX73491	Human nov
4	3143	92.1	3283	3	Aaz36154	AAZ36154	An altern
5	3114.8	91.3	3667	5	Aas70982	AAS70982	DNA encod
6	3114.8	91.3	3667	6	Abs76442	ABS76442	cDNA enco
7	1999.8	58.3	2290	4	Aas26153	AAS26153	Human cDN
8	1999.8	58.3	2290	7	Abx73494	ABX73494	Human nov
9	1979.8	58.0	2010	5	Aas70981	AAS70981	DNA encod
10	1423.8	41.7	1707	7	Aca90176	ACA90176	cDNA enco
11	1096.8	32.1	1186	4	Aas26570	AAS26570	Human cDN
12	1096.8	32.1	1186	4	Abx73911	ABX73911	Human nov
13	776.2	22.7	822	4	Aas26151	AAS26151	Human cDN
14	776.2	22.7	822	7	Abx73492	ABX73492	Human nov
15	712	20.9	1740	6	Aac66035	AAC66035	Human lun
16	712	20.9	1740	6	Abl49254	ABL49254	Human lun
17	712	20.9	1740	6	Abq92440	ABQ92440	Human lun
18	712	20.9	1743	6	Abl49299	ABL49299	Human lun
19	712	20.9	1743	6	Abl49297	ABL49297	Human lun
20	712	20.9	1743	6	Abq92483	ABQ92483	Human lun
21	712	20.9	1743	6	Abq92485	ABQ92485	Human lun
22	712	20.9	1743	8	Ada28540	ADA28540	Recombina
23	712	20.9	1743	8	Ada28537	ADA28537	Recombina
24	710.4	20.8	1740	8		ADA28437	Human lun
25	710.4	20.8	4159	3		AAZ36150	DNA encod
26	710.4	20.8	4181	6		AAC65900	Human lun
27	710.4	20.8	4181	6		ABL65405	Lung canc
28	710.4	20.8	4181	6		ABL49119	Human lun
29	710.4	20.8	4181	8		ABQ92305	Human lun
30	710.4	20.8	4181	8		ADA28265	Human lun
31	710.4	20.8	4181	9		ADD14661	Human src
32	710.4	20.8	4181	8		ADE53470	Human lun
33	710.4	20.8	4434	8		ACH03940	Human cDN
34	710.4	20.8	4601	6		ABA99958	Human KOC
35	709	20.8	1764	6		ABL49283	Human lun
36	709	20.8	1764	6		ABQ92469	Human lun
37	707.4	20.7	1764	2		ADA28518	Recombina
38	690.2	20.2	2224	2		AAZ10617	cDNA enco
39	643.8	18.9	1985	5		AAS76779	DNA encod
40	615.2	18.0	4264	3		AAS86150	DNA encod
41	611	17.9	710	3		AAA02565	Human col
42	570.6	16.7	1946	3		AAZ36153	An altern
43	489.2	14.3	1708	3		AAZ36151	DNA encod
44	469.8	13.8	564	8		AAL57525	Human IGF
45	434	12.7	2780	4		AAK94782	Human ful

ALIGNMENTS

```
RESULT 1
AAZ36152
ID   AAZ36152 standard; DNA; 3412 BP.
XX
AC   AAZ36152;
XX
DT   11-FEB-2000  (first entry)
XX
DE   DNA encoding cancer associated antigen KOC-3.
XX
KW   Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.
XX
OS   Homo sapiens.
XX
PN   WO9954738-A1.
XX
PD   28-OCT-1999.
XX
PF   16-MAR-1999;    99WO-US005766.
XX
PR   17-APR-1998;    98US-00061709.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR   WPI; 2000-013284/01.
XX
PT   Nucleotides representing cancer-associated genes, used to develop
PT   products for the diagnosis, monitoring and treatment of cancers.
XX
PS   Claim 55; Page 41; 44pp; English.
XX
CC   The present sequence represents a cancer associated antigen gene
CC   designated KOC-3. The specification also describes a cancer associated
CC   antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-
CC   37 melanoma cells. The polypeptide has some homology with MAGE-10,
CC   limited to about 210 carboxy terminal amino acids. The amino terminal of
CC   the protein has a repetitive pattern, with repeats rich in serine,
CC   proline, glutamine and leucine, and an almost invariable core of the
CC   peptide given in AAY43877. The CT7 polypeptide can be processed to
CC   peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC   and polypeptides can be used for treating a cancerous condition and
CC   screening for or diagnosing cancerous conditions. The cancer associated
CC   antigens can be used as an immunogenic or vaccine composition with an
CC   adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
```

```
CC   stimulating factor (GM-CSF)
XX
SQ   Sequence 3412 BP; 970 A; 887 C; 804 G; 750 T; 0 U; 1 Other;

     Query Match        100.0%;   Score 3411;   DB 3;   Length 3412;
     Best Local Similarity   100.0%;   Pred. No. 0;
     Matches 3412;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy      1 GGCAGCCGAGGAGGCGAGGAGGCCGGGTACCGGGCCGGGGAGCCGGGCTTCGGGG         60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCAGCCGAGGAGGCGAGGAGGCCGGGTACCGGGCCGGGGAGCCGGGCTTCGGGG         60

Qy     61 AAGAGACGGATGATGAACAAGCTTTACATCGGGGACAGCAGACCTGCCCTGGGGACAGGACCCCGAC        120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGACGGATGATGAACAAGCTTTACATCGGGGACAGCAGACCTGCCCTGGGGACAGGACCCCGAC        120

Qy    121 GACCTCCGGCAGCTCTTTGGGGACAGCAGAAGCTGCCCTGGGGACACAGGTCCTGTGCTGAAG        180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCTCCGGCAGCTCTTTGGGGACAGCAGAAGCTGCCCTGGGGACACAGGTCCTGTGCTGAAG        180

Qy    181 TCCGGCTACGCCTTCGTGGGACTACCCCGACCAGCACTGGGCCCATCGCGCCCATCGAGACC        240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCGGCTACGCCTTCGTGGGACTACCCCGACCAGCACTGGGCCCATCGCGCCCATCGAGACC        240

Qy    241 CTCTCGGGTAAAGTGGAATTGCATCGGGAAAATCATGGAAGTTGGAACAGTCAACACAGAC        300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTCGGGTAAAGTGGAATTGCATCGGGAAAATCATGGAAGTTGGAACAGTCAACACAGAC        300

Qy    301 AAGCTAAGGAGCAGCAGGAAAATTCAGATTCGAACACATCCCTCCTCACCTGCAGTGGGAGGTG        360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCTAAGGAGCAGCAGGAAAATTCAGATTCGAACACATCCCTCCTCACCTGCAGTGGGAGGTG        360

Qy    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAGTCAACACAGAC        420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTGGATGGACTTTTGGCTCAATATGGACAGTGGAGAATGTGGAACAGTCAACACAGAC        420

Qy    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAAAATAGCCATG        480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAAAATAGCCATG        480

Qy    481 GAGAAGCTAAGGCTAAGCGGCATCAGTTGAGAACTACTCTTCAAGATTCCTACATCCCGGAT        540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGAAGCTAAGGCTAAGCGGCATCAGTTGAGAACTACTCTTCAAGATTCCTACATCCCGGAT        540

Qy    541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGCTGGGGACCACTCTTCCCCGG        600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGCTGGGGACCACTCTTCCCCGG        600

Qy    601 GAGCAAGGCCACGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG        660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGCAAGGCCACGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG        660

Qy    661 ATCCTTGGTCCCCACCAGTTTGTTGGTGCCATCATCGGGAAAGAGGGCTTGACCATAAAG        720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCCTTGGTCCCCACCAGTTTGTTGGTGCCATCATCGGGAAAGAGGGCTTGACCATAAAG        720

Qy    721 AACATCACTAAGCAGCAGCCCAGTCCCGGGTAGATATCCATAGAAAGAGAAGAACTCTGGAGCT        780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACATCACTAAGCAGCAGCCCAGTCCCGGGTAGATATCCATAGAAAGAGAAGAACTCTGGAGCT        780

Qy    781 GCAGAGAAGCCTGTCACCATCATCCCACCACCATCCCGAGGGGACTTCTGAAGCATGCCGCATG        840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCAGAGAAGCCTGTCACCATCATCCCACCACCATCCCGAGGGGACTTCTGAAGCATGCCGCATG        840

Qy    841 ATTCTTGAAATCATGCAAACAATGCTTGGTTGGAAGACTGATTGGAAGAAAGGCAGAAATTTG        900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ATTCTTGAAATCATGCAAACAATGCTTGGTTGGAAGACTGATTGGAAGAAAGGCAGAAATTTG        900

Qy    901 AAAATCTTGGCACACAATGCTGATTGGAAGACTGATTGGAAGAAAGGCAGAAATTTG        960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAAATCTTGGCACACAATGCTGATTGGAAGACTGATTGGAAGAAAGGCAGAAATTTG        960

Qy    961 AAGAAAATTGAACATGAACATGAAACACCCGGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC       1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGAAAATTGAACATGAACATGAAACACCCGGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC       1020

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGGCTGTGCCAGTGCT      1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACAACCCGGAAAGAACCATCACTGTGAGGGCACAGTTGAGGGCTGTGCCAGTGCT      1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGGTTGAACCTCAGGGTTGAAAAATGATATGCTGGCTGTTAAC      1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGGTTGAACCTCAGGGTTGAAAAATGATATGCTGGCTGTTAAC      1140

Qy   1141 CAACAAGCCAATCTGATCCCACCAGCAGGAGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGA      1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACAAGCCAATCTGATCCCACCAGCAGGAGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGA      1200

Qy   1201 CTGTCCGTGCTATCTCACCAGCAGGGGGAGTCCCCGCGGGAGTCCCCCGTGCCCCCTACCAC      1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTGCTATCTCACCAGCAGGGGGAGTCCCCGCGGGAGTCCCCCGTGCCCCCTACCAC      1260

Qy   1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGGCCCG      1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCCTTCACTACCCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGGCCCG      1320

Qy   1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG      1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAG      1380

Qy   1381 GCTGTGGGCGCCATCATCGGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGCC      1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GCTGTGGGCGCCATCATCGGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGCC      1440

Qy   1441 GGAGCCTCTATCAAGAATTGCCCCTGCGAAGGCCCAGACGTCAGCGTCAGCGAAAGGATGGTCATC      1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGAGCCTCTATCAAGAATTGCCCCTGCGAAGGCCCAGACGTCAGCGTCAGCGAAAGGATGGTCATC      1500

Qy   1501 ATCACCGGCCACCGAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA      1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATCACCGGCCACCGAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAA      1560

Qy   1561 GAGGAAAACTTCTTAACCCCAAAGAGAAGAAGTGGAAGCGGCCAGATGAAAATGAGAGAAGTG      1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGGAAAACTTCTTAACCCCAAAGAGAAGAAGTGGAAGCGGCCAGATGAAAATGAGAGAAGTG      1620

Qy   1621 TCTTCCACAGCTGGCGCGGGTGATTGGCAAGACCGTGACGAACTGCAGAAC      1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCTTCCACAGCTGGCGCGGGTGATTGGCAAGACCGTGACGAACTGCAGAAC      1680

Qy   1681 TTAACCAGTGCAGAAGTCATCGTCGTGACCAAAACGCCAGATATCGTTCCAACATCCCGAT      1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACCAGTGCAGAAGTCATCGTCGTGACCAAAACGCCAGATATCGTTCCAACATCCCGAT      1740

Qy   1741 ATCGTCGAATTATCGGGCACTCTTTGCTACAGCGGAATCAGGGAA      1800
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATCGTCGAATTATCGGGCACTCTTTGCTACAGCGGAATCAGGGAA      1800

Qy   1801 ATTGTACAACAGTGAAGCAGCAGAAATACCCTCAGGGAGTTCGCCTTCAACAC      1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATTGTACAACAGTGAAGCAGCAGAAATACCCTCAGGGAGTTCGCCTTCAACAC      1860

Qy   1861 AGCAAGTGAGGCCTCCCACCAGCAGCCAGCAAAACAAGGGATGAATGAAACCATCGAG      1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCAAGTGAGGCCTCCCACCAGCAGCCAGCAAAACAAGGGATGAATGAAACCATCGAG      1920

Qy   1921 CTGACAGAATGAGACCAAAGCAGCCAGCAAAGCGGCCAAACCAAGACCATCTGAG      1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGACAGAATGAGACCAAAGCAGCCAGCAAAGCGGCCAAACCAAGACCATCTGAG      1980

Qy   1981 GAATGAGAGTCTGCGGAGGCGGCCAGGGACTCTGCCGGGAACCCCAGGGGC      2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAATGAGAGTCTGCGGAGGCGGCCAGGGACTCTGCCGGGAACCCCAGGGGC      2040

Qy   2041 CGAGGAGGGCGGGAAGGTCAGTTGCCGAGACAATAACAATCTCATCTTTGCAGGATTTGAGC      2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  3181 GTGTTTTTACCTTGCACCTTGCTCTTGCTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 GTGTTTTTACCTTGCACCTTGCTCTTGCTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240

Qy  3241 TTGGAGCATTTTTTATTTTTTAATAAATGAGTTGGAAAAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTTATTTTTTAATAAATGAGTTGGAAAAAAAATAAGATATCAACT 3300

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGTGTTCTGAATTGTCTTCCCCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGTGTTCTGAATTGTCTTCCCGCT 3360

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2

AAS26150

```
ID   AAS26150 standard; cDNA; 3694 BP.
XX
AC   AAS26150;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 329.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0186664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0220664P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
```

```
Db  2041 CGAGGAGGGGGGCGGGGAAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCGCCTCCCGCC 2100

Qy  2101 CCCCAGGGCTTCTGCAGGCTTCAGCCATTCACCTTCACCTTCAAAGCCAAGC 2160
         |||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGGCTTCAGCCATTCACCTTCACCTTCAAAGCCAAGC 2160

Qy  2161 CTCCCACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         |||||||||||||||||||||||||||||||||||
Db  2161 CTCCCACGACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2280
         |||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGACGCGGTAATATAT 2340
         |||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGACGCGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATATCAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
         |||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATATCAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTAAAGAGAAAGCAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         |||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTAAAGAGAAAGCAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGACAAAATTCACCCAGAGGGGAAAT 2520
         |||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCACACAAAATTCACCCAGAGGGGAAAT 2520

Qy  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGGATA 2580
         |||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGAGGGATA 2580

Qy  2581 CCGTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGCTCTAGCTCATCACACCATT 2640
         |||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGCTCTAGCTCATCACACCATT 2640

Qy  2641 TCTCTTGCTTCAGAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         |||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTTGCTTCACAGGTTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700

Qy  2701 CTCTCTGTTATCTCCCCTCCCCCCTTCTTCTCCCATCCCACCCCCCCCCCCCAGGCAAAGCA 2760
         |||||||||||||||||||||||||||||||||||
Db  2701 CTCTCTGTTATCTCCTCCATCTCCATTCTTCTCCCATCTCCATTCTTCTCCCATCCCATCTTTTGAA 2760

Qy  2761 TTTTCCTCATCCTCCAATCTCAATCCCGTATCTCACGCACCCCCCCCCCCCAGGCAAAGCA 2820
         |||||||||||||||||||||||||||||||||||
Db  2761 TTTTCCTCATCCTCCAATCTCAATCCCGTATCTCACGCACCCCCCCCCCCCAGGCAAAGCA 2820

Qy  2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAA 2880
         |||||||||||||||||||||||||||||||||||
Db  2821 GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCCTCAA 2880

Qy  2881 CTTACACTTGGTTACTTCAAAAGAACAACCAACCAATCAACAAAGAAAAATTCCACAATG 2940
         |||||||||||||||||||||||||||||||||||
Db  2881 CTTACACTTGGTTACTTCAAAAGAACAACCAACCAATCAACAAAGAAAAATTCCACAATG 2940

Qy  2941 AGGAAAACAGGAACCCACCACCAATCAACCAATCAACAAAGAAAAATTCCACAATG 3000
         |||||||||||||||||||||||||||||||||||
Db  2941 AGGAAAACAGGAACCCACCACCAATCAACCAATCAACAAAGAAAAATTCCACAATG 3000

Qy  3001 AAAGAATGTATTTGTCTTTTGTCTTTTGCATTTTGGTTATAAGCCATCAAATTCAGCAAAATG 3060
         |||||||||||||||||||||||||||||||||||
Db  3001 AAAGAATGTATTTGTCTTTTGTCTTTTGCATTTTGGTTATAAGCCATCAAATTCAGCAAAATG 3060

Qy  3061 ATTCCTTTCTTTAAAAAATGTGAGAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGGC 3120
         |||||||||||||||||||||||||||||||||||
Db  3061 ATTCCTTTCTTTAAAAAATGTGAGAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120

Qy  3121 CAGGGGCTTAAATTCACAGATTTTTTTAACGGAGAAAAACACACAGAGGAAGCTACCCTCAG 3180
         |||||||||||||||||||||||||||||||||||
Db  3121 CAGGGGCGTAAATTCACAGATTTTTTTTAACGGAGAAAAACACACAGAGGAAGCTACCCTCAG 3180
```

```
PR   14-AUG-2000;   2000US-0225757P.          PR   08-NOV-2000;   2000US-0246609P.
PR   14-AUG-2000;   2000US-0225758P.          PR   08-NOV-2000;   2000US-0246610P.
PR   14-AUG-2000;   2000US-0225759P.          PR   08-NOV-2000;   2000US-0246611P.
PR   18-AUG-2000;   2000US-0226279P.          PR   08-NOV-2000;   2000US-0246613P.
PR   22-AUG-2000;   2000US-0226681P.          PR   17-NOV-2000;   2000US-0249207P.
PR   22-AUG-2000;   2000US-0226868P.          PR   17-NOV-2000;   2000US-0249208P.
PR   22-AUG-2000;   2000US-0227182P.          PR   17-NOV-2000;   2000US-0249209P.
PR   23-AUG-2000;   2000US-0227009P.          PR   17-NOV-2000;   2000US-0249210P.
PR   30-AUG-2000;   2000US-0228924P.          PR   17-NOV-2000;   2000US-0249211P.
PR   01-SEP-2000;   2000US-0229287P.          PR   17-NOV-2000;   2000US-0249212P.
PR   01-SEP-2000;   2000US-0229343P.          PR   17-NOV-2000;   2000US-0249213P.
PR   01-SEP-2000;   2000US-0229344P.          PR   17-NOV-2000;   2000US-0249214P.
PR   01-SEP-2000;   2000US-0229345P.          PR   17-NOV-2000;   2000US-0249215P.
PR   05-SEP-2000;   2000US-0229509P.          PR   17-NOV-2000;   2000US-0249216P.
PR   05-SEP-2000;   2000US-0229513P.          PR   17-NOV-2000;   2000US-0249217P.
PR   06-SEP-2000;   2000US-0230437P.          PR   17-NOV-2000;   2000US-0249218P.
PR   06-SEP-2000;   2000US-0230438P.          PR   17-NOV-2000;   2000US-0249244P.
PR   08-SEP-2000;   2000US-0231242P.          PR   17-NOV-2000;   2000US-0249245P.
PR   08-SEP-2000;   2000US-0231243P.          PR   17-NOV-2000;   2000US-0249264P.
PR   08-SEP-2000;   2000US-0231244P.          PR   17-NOV-2000;   2000US-0249265P.
PR   08-SEP-2000;   2000US-0231413P.          PR   17-NOV-2000;   2000US-0249297P.
PR   08-SEP-2000;   2000US-0231414P.          PR   17-NOV-2000;   2000US-0249299P.
PR   08-SEP-2000;   2000US-0232080P.          PR   17-NOV-2000;   2000US-0249300P.
PR   08-SEP-2000;   2000US-0232081P.          PR   01-DEC-2000;   2000US-0250160P.
PR   12-SEP-2000;   2000US-0231968P.          PR   01-DEC-2000;   2000US-0250391P.
PR   14-SEP-2000;   2000US-0232397P.          PR   05-DEC-2000;   2000US-0251030P.
PR   14-SEP-2000;   2000US-0232398P.          PR   05-DEC-2000;   2000US-0251198P.
PR   14-SEP-2000;   2000US-0232399P.          PR   05-DEC-2000;   2000US-0256719P.
PR   14-SEP-2000;   2000US-0232400P.          PR   06-DEC-2000;   2000US-0251479P.
PR   14-SEP-2000;   2000US-0232401P.          PR   08-DEC-2000;   2000US-0251856P.
PR   14-SEP-2000;   2000US-0233063P.          PR   08-DEC-2000;   2000US-0251868P.
PR   14-SEP-2000;   2000US-0233064P.          PR   08-DEC-2000;   2000US-0251869P.
PR   14-SEP-2000;   2000US-0233065P.          PR   08-DEC-2000;   2000US-0251989P.
PR   21-SEP-2000;   2000US-0234223P.          PR   08-DEC-2000;   2000US-0251990P.
PR   21-SEP-2000;   2000US-0234274P.          PR   11-DEC-2000;   2000US-0254097P.
PR   25-SEP-2000;   2000US-0234997P.          PR   05-JAN-2001;   2001US-0259678P.
PR   25-SEP-2000;   2000US-0234998P.          XX
PR   26-SEP-2000;   2000US-0235484P.          PA   (HUMA-) HUMAN GENOME SCI INC.
PR   27-SEP-2000;   2000US-0235834P.          XX
PR   27-SEP-2000;   2000US-0235836P.          PI   Rosen CA,  Barash SC,  Ruben SM;
PR   29-SEP-2000;   2000US-0236327P.          XX
PR   29-SEP-2000;   2000US-0236367P.          DR   WPI; 2001-488783/53.
PR   29-SEP-2000;   2000US-0236368P.          DR   P-PSDB; AAU16163.
PR   29-SEP-2000;   2000US-0236369P.          XX
PR   29-SEP-2000;   2000US-0236370P.          PT   New nucleic acid molecules encoding 461 human secreted proteins for
PR   02-OCT-2000;   2000US-0236802P.          PT   diagnosing, preventing, treating or ameliorating medical conditions and
PR   02-OCT-2000;   2000US-0237037P.          PT   used as food additives or preservatives.
PR   02-OCT-2000;   2000US-0237038P.          XX
PR   02-OCT-2000;   2000US-0237039P.          PS   Claim 1; SEQ ID NO 329; 980pp; English.
PR   02-OCT-2000;   2000US-0237040P.          XX
PR   13-OCT-2000;   2000US-0239935P.          CC   The invention relates to isolated nucleic acid molecules and their
PR   13-OCT-2000;   2000US-0239937P.          CC   encoded secreted proteins. The nucleic acids and proteins are used to
PR   20-OCT-2000;   2000US-0240960P.          CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
PR   20-OCT-2000;   2000US-0241221P.          CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
PR   20-OCT-2000;   2000US-0241785P.          CC   in diagnosing a pathological condition or susceptibility to a
PR   20-OCT-2000;   2000US-0241786P.          CC   pathological condition. Antibodies to the proteins can also be used in
PR   20-OCT-2000;   2000US-0241787P.          CC   alleviating symptoms associated with the disorders and in diagnostic
PR   20-OCT-2000;   2000US-0241808P.          CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
PR   20-OCT-2000;   2000US-0241809P.          CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
PR   01-NOV-2000;   2000US-0241826P.          CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
PR   01-NOV-2000;   2000US-0244617P.          CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
PR   08-NOV-2000;   2000US-0246474P.          CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
PR   08-NOV-2000;   2000US-0246475P.          CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
PR   08-NOV-2000;   2000US-0246476P.          CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
PR   08-NOV-2000;   2000US-0246477P.          CC   and many other disorders listed in the specification. The polypeptides
PR   08-NOV-2000;   2000US-0246478P.          CC   can also be used to aid wound healing and epithelial cell proliferation,
PR   08-NOV-2000;   2000US-0246523P.          CC   to prevent skin aging due to sunburn, to maintain organs before
PR   08-NOV-2000;   2000US-0246524P.          CC   transplantation, for supporting cell culture of primary tissues, to
PR   08-NOV-2000;   2000US-0246525P.          CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
PR   08-NOV-2000;   2000US-0246526P.          CC   as a food additive or preservative to increase or decrease storage
PR   08-NOV-2000;   2000US-0246527P.          CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
PR   08-NOV-2000;   2000US-0246528P.          CC   minerals, cofactors and other nutritional components. The present
PR   08-NOV-2000;   2000US-0246532P.          CC   sequence encodes a novel secreted protein of the invention. Note: The
```

CC   sequence data for this patent did not form part of the printed

Query Match   97.3%;   Score 3320.6;   DB 4;   Length 3694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 3343;   Conservative   1;   Mismatches   1;   Indels   3;   Gaps   2;

```
Qy   68 GGATGATGAACAAGCTTACATCGGGAACCTGAGCCCCGCCGTCACCGACGACCTCC   127
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   91 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGTCACCGACGACCTCC  150

Qy  128 GGCAGCTCTTTGGGACAGGAAGAGCTGCCCTGGGCGGGACAGGTCCTGTGAAGTCCGGCT  187
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 GGCAGCTCTTTGGGACAGGAAGAGCTGCCCTGGGCGGGACAGGTCCTGTGAAGTCCGGCT  210

Qy  188 ACGGCCTTCGTGGACTACCCCGACCACCCGAGAACTGGGCCATCGCGCGACCCCTCTCGG  247
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 ACGGCCTTCGTGGACTACCCCGACCACCCGAGAACTGGGCCATCGCGCGACCCCTCTCGG  270

Qy  248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTGAAGTTGATTACTCAGTCTTAAAAGCTAA  307
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTGAAGTTGATTACTCAGTCTTAAAAGCTAA  330

Qy  308 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGGATG  367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGGATG  390

Qy  368 GACTTTGGCTCAATATGGGACAGTGGAAATGTGAACAAGTCAACACAGACACAGAAA  427
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391 GACTTTGGCTCAATATGGGACAGTGGAAATGTGAACAAGTCAACACAGACACAGAAA  450

Qy  428 CCGCCGTTGTCAACCTCACATATGCAACAGAGAAGCAAAAATAGCCATGGAAGAGC  487
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451 CCGCCGTTGTCAACGTCAACATATGCAACAGAGAAGCAAAAATAGCCATGGAAGAGC  510

Qy  488 TAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGG  547
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  511 TAAGGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGG  570

Qy  548 TGAGTCCCCTTCGCCCCCCAGCGGCCCACTCTTCCCGGGACCACTCTTCCCGGGAGCAAG  607
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  571 TGAGTCCCCTTCGCCCCCCAGCGGCCCAGCGTGGGGACCACTCTTCCCGGGAGCAAG  630

Qy  608 GCCAGCGCCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGGCTGCGGATCCTG  667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  631 GCCAGCGCCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGGCTGCGGATCCTG  690

Qy  668 TCCCCACCCAGTTGTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAACATCA  727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  691 TCCCCACCCAGTTGTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAACATCA  750

Qy  728 CTAAGCAGACCAGCCAGTCCCGGTAGATATCGGAAAGGAGCTGCAGAGA  787
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  751 CTAAGCAGACCAGCCAGTCCCGGTAGATATCGGAAAGGAGCTGCAGAGA  810

Qy  788 AGCCTTGTCACCATCCATGCCACCCACCCACCCGAGAGGGGGACTTCTGAAGCATGCCGCATGATTCTTG  847
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  811 AGCCTTGTCACCATCCATGCCACCCACCCACCCGAGAGGGGGACTTCTGAAGCATGCCGCATGATTCTTG  870

Qy  848 AAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGATTCCTCTGAAAATCT  907
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  871 AAATCATGCAGAAAGAGGCAGATGACCAAACTAGCCGAAGATTCCTCTGAAAATCT  930

Qy  908 TGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAGGGCAGAAATTGAAGAAAA  967
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  931 TGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGTGAAAAGAGGGCAGAAATTTGAGAGAAA  990

Qy  968 TTGAACAATGAAACAGGAGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA  1027
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  991 TTGAACAATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA  1050

Qy 1028 ACCCGGAAAAGAACCATCACTGTGAGAGGCAGTTCAGGCCTGTGCCAGTCCAGTCGAGATAG  1087
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1051 ACCCGGAAAAGAACCATCACTGTGAGAGGCAGTTCAGGCCTGTGCCAGTCCAGTCGAGATAG  1110
```

```
Qy 1088 AGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATAAATGATGATGTGGCTGTTAACCAACAAG  1147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1111 AGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATAAATGATGATGTGGCTGTTAACCAACAAG  1170

Qy 1148 CCAATCTGATCCCAGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCG  1207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1171 CCAATCTCATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTGTCCG  1230

Qy 1208 TGCTATCTCCACCAGCAGGGGCCCCCGGAGCTCCTGCTGTGAAGTCCGGCT  1267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1231 TGCTATCTCCACCAGCAGGGGCCCCCGCTGCCCCCTGCTGTGAAGTCCGGCT  1290

Qy 1268 CTACCCACTCCGGATACTTCTCCAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGG  1327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1291 CTACCCACTCCGGATACTTCTCCAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGCTGTGG  1350

Qy 1328 ATCATCACTCTTATCCAGAGCAGGAGATTCTGGCGAGATCTGCCGGAGCCT  1387
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1351 ATCATCACTCTTATCCAGAGCAGGAGATTCTGGCGAGATTCGCCGGAGCCT  1410

Qy 1388 GCGGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCT  1447
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1411 GCGGCCATCATCGGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGAGCCT  1470

Qy 1448 CTATCAAGAATTGCCCCTGCGAAGGCCCAGAGCTCAGCGAGGATGGTCATCATCACCACG  1507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1471 CTATCAAGAATTGCCCCTGCGAAGGCCCAGAGCTCAGCGAGGATGGTCATCATCACCACG  1530

Qy 1508 GGCACCCGAAGCCCAGTTCAAGGCCCAGGACGGATCTTGGGAGAAACTGAAAGAGAGAAA  1567
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1531 GGCACCCGAAGCCCAGTTCAAGGCCCAGGACGGATCTTGGGAGAAACTGAAAGAGAGAAA  1590

Qy 1568 ACTTCTTTAACCCAAAGAGAAGTGAAGCTGGAAGTCAGATCAGATCAGAGTGCCCTTCCA  1627
         ||||||||||||||||||||||||||||||||:|||::|||||||||||||||||||||
Db 1591 ACTTCTTTAACCCAAAGAGAAGTGAAGCTGAASCGGCCATATCAGATCAGAGTGCCCTTCCA  1650

Qy 1628 CAGCTGGCGCGGGTGATTGGCAAAGGTGGCAAGACCGTGAAGAACTGAGAACTTAACCA  1687
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1651 CAGCTGGCGCGGGTGATTGGCAAAGGTGGCAAGACCGTGAAGAACTGAGAACTTAACCA  1710

Qy 1688 GTGCAGAGTCATCGTGCTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA  1747
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1711 GTGCAGAGTCATCGTGCTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGATCGTCA  1770

Qy 1748 GAATTATCGGGCACTTCTTTGCTAGCCAGAATACCCTGGAAGTCAGGGAAATTGTAC  1807
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1771 GAATTATCGGGCACTTCTTTGCTAGCCAGAATACCCTGGAAGTCAGGGAAATTGTAC  1830

Qy 1808 AACAGGTGAAGCAGCAGCAGGACCAGAACACAGACAAAACCGATCGGGAGAACCAAACAAGT  1867
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1831 AACAGGTGAAGCAGCAGCAGGACCAGAACAACAGACAAAACCGATCGGGAGCAAGT  1890

Qy 1868 GAGGCTCCCACAGAGATGAATGGAATGCGATGAATGTAGCCCTTCCAACACCTGACAG  1927
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1891 GAGGCTCCCACAGCGATGAATGGAATGCGATGAATGTAGCCCTTCCAACACCTGACAG  1950

Qy 1928 AATGAGACCAAACGACCAGCAGCGATGGGAGCAAACCAAACCATCTGAGGAATAG  1987
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1951 AATGAGACCAAACGACCAGCAGCGATCGGGAGCAAACCAAACCATCTGAGGAATGAG  2010

Qy 1988 AAGTCTGCGGGAGGCTCAGGTTTGCCAGAACCCAGGGACCCCCGGCGCTCCGGAGGAG  2047
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2011 AAGTCTGCGGGAGGCTCAGGTTTGCCAGAACCCAGGGACCCCCGCCCCTCCGGAGGAG  2070

Qy 2048 GGGCGGGAGGTCAGCGAGGTTTGCCAGAACACCCGGGCTCCCGCCCCAGG  2107
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 2071 GGGCGGGGAAGGTCAGCGAGGTTTGCCAGAACACCCGGGCTCCCGCCCCCAGG  2130

Qy 2108 GCTTCTGCGAGGCTTCAGCGCATCCACCATCCACTCTCCTGAACTCCCAC  2167
         |||||||||||||||||||||||||||||||||||||||||:|||||||
Db 2131 GCTTCTGCGAGGCTTCAGCGCATCCACCATCCACTCCACTCTCTCGAACTCCCAC  2190
```

```
Qy   2168  GACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTCAAGCCAAGCAAAATGC  2227
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2191  GACGCTATCCCTTTAGTGAACTAACATAGGTGAACGTGTCAAGCCAAGCAAAATGC  2250

Qy   2228  ACACCCTTTTCTGTGGCAAATCGTCTCGTACATGTGTGTACATATTAGAAGGGAAGA  2287
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2251  ACACCCTTTTCTGTGTGCAAATCGTCTCGTACATGTGTGTACATATTAGAAGGGAAGA  2310

Qy   2288  TGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTAGAA  2347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311  TGTTAAGAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTAGAA  2370

Qy   2348  ATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTATTTTTTTTTCT  2407
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2371  ATAATATATCAAATAACTAACTCAACTACTCCAATTTTAATCAATTATTATTTAATTTTTTTTTCT  2430

Qy   2408  TTTTAAGAGAAAGCAGAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  2467
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2431  TTTTAAGAGAGAAAGCAGAGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC  2490

Qy   2468  GGTGTAGAGAGGAGCTTTGAGGCACCCCGCACAAAATTCACCCGAGAGGGAAATCTCGTCG  2527
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2491  GGTGTAGAGAGGAGCTTTGAGGCACCCCGCACAAAATTCACCCGAGAGGGAAATCTCGTCG  2550

Qy   2528  GAAGGACACTCACGGCAGTTCTCGGATCACCTGTATGTCAACAGAGGGATACCGTCTC  2587
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2551  GAAGGACACTCACGGCAGTTCTCGGATCACCTGTGTATGTCAACAGAGGGATACCGTCTC  2610

Qy   2588  CTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATACACCCATTTCTCTTT  2647
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2611  CTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATACACCCATTTCTCTTT  2670

Qy   2648  GCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCCTCTCTG  2707
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2671  GCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCTG  2730

Qy   2708  TTTATCTCTCCCCTCCCCTCCCCTCCCCTCCCCTCCATCTCCATTCTTTGAATTCCTC  2767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2731  TTTATCTCTCCCCTCCCCTCCCCTCCCCTCCATCTCCATTCTTTGAATTCCTC  2790

Qy   2768  ATCCCTCCATCTCAATCCGTATCTACGGCA--CCCCCCCCCAGGCAAAGCAGTTGCT  2825
           |||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db   2791  ATCCCTCCATCTCAATCCGTATCTACGGCACCCCGTACCCCCCCCCCAGGCAAAGCAGTTGCT  2850

Qy   2826  CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACAACACCAGCCTCAACTTAC  2885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2851  CTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACAACACCAGCCTCAACTTAC  2910

Qy   2886  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2945
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2911  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA  2970

Qy   2946  AACAGGAACCCACCACCAAAACCCAAATCCACCCAACAAAGAAAAAATTCCACAATGAAAGA  3005
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2971  AACAGGAACCCACCACCAAAACCCAAATCCACCCAACAAAGAAAAAATTCCACAATGAAAGA  3030

Qy   3006  ATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCGAAAATGATTCC  3065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3031  ATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCGAAAATGATTCC  3090

Qy   3066  TTTCTTT-AAAAAAAAAAAAAAATGTGGAGGAAGTGGAGGAAGTAGAAATTTACCAAGGTGTTGGCCCAGG  3124
           ||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db   3091  TTTCTTTAAAAAAAAAAAAAAAATGTGGAGGAAGTGGAGGAAGTAGAAATTTACCAAGGTGTTGGCCCAGG  3150

Qy   3125  GCGTTAAATTCACAGATTTTTTCTTGCTCTGTGTTTCCCTTGAGATTTTGTATAAGCTACCTCAGTGT  3184
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3151  GCGTTAAATTCACAGATTTTTTCTTGCTCTGTGTTTCCCTTGAGATTTTGTATAAGCTACCTCAGTGT  3210

Qy   3185  TTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTGAGATTTTGTATAAGCTGATAGTTGG  3244
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3211  TTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTGAGATTTTGTATAAGCTGATAGTTGG  3270

Qy   3245  AGCATTTTTTATTTTTTTAATAAAATGAGTTGGAAAAAATAAGATATCAACTGCCA  3304
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3271  AGCATTTTTTATTTTTTTAATAAAATGAGTTGGAAAAAATAAGATATCAACTGCCA  3330

Qy   3305  GCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGTGTCTTCTGAATTGTCTTCCGCTAGCC  3364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3331  GCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGTGTCTTCTGAATTGTCTTCCGCTAGCC  3390

Qy   3365  AAGAACCNATATGGCCTTCTTTTGGACAACCTTGAAAATGTTTATTT  3412
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   3391  AAGAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3438
```

```
RESULT 3
ABX73491
ID   ABX73491 standard; DNA; 3694 BP.
XX
AC   ABX73491;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #319.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US200213753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180528P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0229924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
```

```
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   20-OCT-2000;   2000US-0240940P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241809P.
PR   01-NOV-2000;   2000US-0244617P.
PR   17-NOV-2000;   2000US-0249299P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55231.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 329; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
XX   human novel polynucleotides of the invention
SQ   Sequence 3694 BP; 1103 A; 934 C; 839 G; 817 T; 0 U; 1 Other;
```

```
Query Match             97.3%;   Score 3320.6;   DB 7;   Length 3694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 3343;  Conservative   1;  Mismatches   1;  Indels   3;  Gaps   2;

Qy     68 GGATGATGAACAAGCTTTACATCGGGAACCTGGAACCCGGCCCCGTCACGCCGACGACCTCC 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 GGATGATGAACAAGCTTTACATCGGGAACCTGGAACCCGGCCCCGTCACGCCGACGACCTCC 150

Qy    128 GGCAGGCTCTTTGGGGACAGGAAGCTGCCCCTGGCCCCTGGCGGGACAGAGTCCTGCTGAAGTCCGGCT 187
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 GGCAGGCTCTTTGGGGACAGGAAGCTGCCCCTGGCCCCTGGCGGGACAGAGTCCTGCTGAAGTCCGGCT 210

Qy    188 ACGCCTTCGTGACTACCCCGACCAGAACTGGGACATCCCGCCGCCATCCGGCCAGACCCTCTCGG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 ACGCCTTCGTGACTACCCCGACCAGAACTGGGACATCCCGCCGCCATCCGGCCAGACCCTCTCGG 270

Qy    248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTTAAAAAGCTAA 307
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTTCTTAAAAAGCTAA 330

Qy    308 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCACTGCAGTGGAGTGTTGGATG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    331 GGAGCAGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGTGTTGGATG 390
```

```
Qy    368 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGA 427
          |||||||||||||||||||||||||||||||||||
Db    391 GACTTTTGGCTCAATATGGACAGTGGAGAATGTGGA 450

Qy    428 CCGCCGTTGTCAACGTCACATATGCAACAGAGAGAAGC 487
          ||||||||||||||||||||||||||||||||||||||
Db    451 CCGCCGTTGTCAACGTCACATATGCAACAGAGAGAAGC 510

Qy    488 TAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAT 547
          |||||||||||||||||||||||||||||||||||||
Db    511 TAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAT 570

Qy    548 TGAGCTCCCCTTGCCCCCTCAGCGAGCCCAGCCCAGCCG 607
          |||||||||||||||||||||||||||||||||||||||
Db    571 TGAGCTCCCCTTGCCCCCTCAGCGAGCCCAGCCCAGCCG 630

Qy    608 GCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGGTGCGGATCCTGG 667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    631 GCCACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGGCTGCGGATCCTGG 690

Qy    668 TCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGAGGAGGGCTTGACCATAAAGAACATCA 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    691 TCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGAGGGCTTGACCATAAAGAACATCA 750

Qy    728 CTAAGCAGACCCAGTCCCGGGTAGAATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGA 787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    751 CTAAGCAGACCCAGTCCCGGGTAGAATATCCATAGAAAAGAGAACTCTGGAGCTGCAGAGA 810

Qy    788 AGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    811 AGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 870

Qy    848 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCCAAGAGATTCCTCTGAAATCT 907
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    871 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCCAAGAGATTCCTCTGAAAATCT 930

Qy    908 TGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCCAGAAATTTGAAGAAAA 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    931 TGGCACAAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCCAGAAATTTGAAGAAAA 990

Qy    968 TTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1027
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    991 TTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1050

Qy   1028 ACCCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCTGTGCCAGTGTGCTGAGATAG 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1051 ACCCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCTGTGCCAGTGTGCTGAGATAG 1110

Qy   1088 AGATTATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1111 AGATTATGAAGAAGCTGCGTGAGGGCTTTGAAAATGATATGCTGGCTGTTAACCAACAAG 1170

Qy   1148 CCAATCTGATCCCGGGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGACTGTCCG 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CCAATCTGATCCCGGGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGACTGTCCG 1230

Qy   1208 TGCTATCTCCACCAGCAGGGGCCCCGGGAGCTCCCCCGCTGCCCCCGCTGCACCACCCCTTCA 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1231 TGCTATCTCCACCAGCAGGGGCCCCGGGAGCTCCCCCGCTGCCCCCGCTGCACCACCCCTTCA 1290

Qy   1268 CTACCCACTCCGGATACTTCTCGTACCCCTGTACCCGCCAGTTTGGCCGTTCCCGC 1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 CTACCCACTCCGGATACTTCTCCAGCCTGTACCCGCCAGTTTGGCCGTTCCCGC 1350

Qy   1328 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTCAGTCTTCAGTCTTCAGGGTGTGG 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 ATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCAGTCTTCAGTCTTCAGGGTGTGG 1410

Qy   1388 GGCCCATCATCCGGAAGAAGAGGGGCCACACATCAAACAGCTGGCCGAGATTCGCCCGGAGCCT 1447
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1411 GCGCCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCCGAGATTGCCCGGAGCCT 1470
```

```
Qy   1448  CTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGATGGTCATCATCACCG  1507
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471  CTATCAAGAATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGATGGTCATCATCACCG  1530

Qy   1508  GGCCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAA   1567
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1531  GGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAGGAAA  1590

Qy   1568  ACTTCTTTAACCCCAAAGAAGAAGTGAAGCGCATATCAGAGTCGTCCCTCTTCCA      1627
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1591  ACTTCTTTAACCCCAAAGAAGAAGTGAAGCGCATATCAGAGTCGTCCCTCTTCCA      1650

Qy   1628  CAGCTGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTTAACCA        1687
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1651  CAGCTGCGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTTAACCA        1710

Qy   1688  GTGCAGAAGTCATCGTGCCCTGTGACCAAGCGCCAGATGAAAGTGAAGTGATCGTCA     1747
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1711  GTGCAGAAGTCATCGTGCCCTGTGACCAAGCGCCAGATGAAAGTGAAGTGATCGTCA     1770

Qy   1748  GAATTATCGGGGCACTTCTTTGCTCAGCCAGACTGCACAGCGCCAGATCAGGGAAATTGTAC  1807
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1771  GAATTATCGGGGCACTTCTTTGCTCAGCCAGACTGCACAGCGCCAGATCAGGGAAATTGTAC  1830

Qy   1808  AACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGT    1867
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1831  AACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGT    1890

Qy   1868  GAGGCTCCCACAGGCCACCAGGCAAAACAACGATGAAGTGAGCCTTCCAACACCTGACAG   1927
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1891  GAGGCTCCCACAGGCCACCAGGCAAAACAACGATGAAGTGAGCCTTCCAACACCTGACAG   1950

Qy   1928  AATGAGACCAAACGCAGCGAGCCAGTCCAGCAAAACCAAGACCATCTGAGGAATAGAG     1987
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1951  AATGAGACCAAACGCAGCGAGCCAGTCCAGCAAAACCAAGACCATCTGAGGAATAGAG     2010

Qy   1988  AAGTCTGCGGAGGCGGCGCGCCCAGGGACTCTGCCGAGGACTTCGCCGAGGCCCGAGGAG    2047
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2011  AAGTCTGCGGAGGCGGCGCGCCCAGGGACTCTGCCGAGGACTTCGCCGAGGCCCGAGGAG    2070

Qy   2048  GGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGGGCCTCCCGCCCCCGGG         2107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2071  GGGCGGGGAAGGTCAGCCAGGTTGCCAGAACCACCGGGCCTCCCGCCCCCGGG         2130

Qy   2108  GCTTCTGCGACCTTCAGCGCTTCACCATCCACTTCACCCTGAAATCTCCTGAACTCCCAC    2167
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2131  GCTTCTGCGACCTTCAGCGCTTCACCATCCACTTCACCCTGAAATCTCCTGAACTCCCAC    2190

Qy   2168  GACGCTATCCCTTTTAGTTAGTAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGC  2227
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2191  GACGCTATCCCTTTTAGTTAGTAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAATGC  2250

Qy   2228  ACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAAGGGAAGA   2287
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2251  ACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAAGGGAAGA   2310

Qy   2288  TGTTAAGATATGTGCCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTTAGAA    2347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311  TGTTAAGATATGTGCCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATATTTTAGAA    2370

Qy   2348  ATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTTTTCT    2407
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2371  ATAATATATCAAATAACTAACTCCAATTTTAATCAATTATTAATTTTTTTTTTTTTCT    2430

Qy   2408  TTTTAAAGAGAAAGCAGCAGCCTTTCTTAGACTTTTAAAGAAATAAAGTCTTTGGGAGGTCTCAC  2467
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2431  TTTTAAAGAGAAAGCAGCAGCCTTTCTTAGACTTTTAAAGAAATAAAGTCTTTGGGAGGTCTCAC  2490

Qy   2468  GGTGTAGAGGAGCTTTGAGGCCACCCGCACCCCACCAAATTCACCCAGAGGGAAATCTCGTCG  2527
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2491  GGTGTAGAGGAGCTTTGAGGCCACCCGCACCCCACCAAATTCACCCAGAGGGAAATCTCGTCG  2550

Qy   2528  GAAGGACACTCACCGCAGTTCTGATCACCTGTGATCGTCTC                        2587
```

```
Db   2551  GAAGGACACTCACCGCAGTTCTGATCACCTGTGATCGTCTCTGCCTGTGATGTCAACAGAGAGGGATACCGTCTC  2610

Qy   2588  CTTGAAGAGGAAACTCTGTCACTCCTCAGCCTGCTTCATCCTCAGCTCATCATACACCCATTTCTCTTT         2647
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   2611  CTTGAAGAGGAAACTCTGTCACTCCTCAGCCTGCTTCATCCTCAGCTCATATATAATTCTCGTCTCTCTTT      2670

Qy   2648  GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCCTG              2707
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2671  GCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCTCCTG              2730

Qy   2708  TTTATCTCTCCCCTCCCCTCCCCTTCTTCCATCTCCATCTCCTTTGAATTCCTC                 2767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2731  TTTATCTCTCCCCTCCCCTCCCCTTCTTCCATCTCCATCTCCTTTGAATTCCTC                 2790

Qy   2768  ATCCCTCCATCTCAATCCCGTATCTACGCA--CCCCCCCCCCCCCCCAGGCAAAGCAGTGCT          2825
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db   2791  ATCCCTCCATCTCAATCCCGTATCTACGCACCCCCCCCCCCCCCCCCCAGGCAAAGCAGTGCT        2850

Qy   2826  CTGAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAAACCAGCCTCCAACTTAC          2885
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2851  CTGAGTATCACATCACACAAAAGGAACAAAAGGCGAAACACACAAACCAGCCTCCAACTTAC          2910

Qy   2886  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA           2945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2911  ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAGGAA           2970

Qy   2946  AACAGGAACCCACCAACCAACCAATCAACAAGAAGAAAAATTCCACAATGAAAGA                3005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2971  AACAGGAACCCACCAACCAACCAATCAACAAGAAGAAAAATTCCACAATGAAAGA                3030

Qy   3006  ATGTATTTTGTCTTTTGTCATTTGGTGTGTATAAGCCATCAATCAATATTCAGCAAAATGATTCC      3065
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3031  ATGTATTTTGTCTTTTGTCATTTGGTGTGTATAAGCCATCAATCAATATTCAGCAAAATGATTCC      3090

Qy   3066  TTTCTTT-AAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGTTGTTGGCCCAGG             3124
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3091  TTTCTTTTAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGTTGTTGGCCCAGG           3150

Qy   3125  GCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAGCTACCTCAGGTGT          3184
           ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   3151  GCGTTAAATTCACAGATTTTTTTTAACGAGAAAAACACACAGAGAGCTACCTCAGTGT         3210

Qy   3185  TTTTACCTCAGCACCTTGCTCCTTGTGTTCCCTTAGAGATTTGTAAAGCTTGATAGTTGG         3244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3211  TTTTACCTCAGCACCTTGCTCCTTGTGTTCCCTTAGAGATTTGTAAAGCTTGATAGTTGG         3270

Qy   3245  AGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA              3304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3271  AGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTGCCA             3330

Qy   3305  GCCTGGAGAAGGTGACAGTGCAAGTGTGCAACAGCTGTTCTGAATTGCCTAGCC                3364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3331  GCCTGGAGAAGGTGACAGTGCCAAGTGTGCAACAGCTGTTCTGAATTGCCTAGCC               3390

Qy   3365  AAGAACCNATATGGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT                      3412
           ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3391  AAGAACCTATATGGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT                      3438
```

```
RESULT 4

AAZ36154
ID   AAZ36154 standard; DNA; 3283 BP.
XX
AC   AAZ36154;
XX
DT   11-FEB-2000 (first entry)
XX
DE   An alternative form of DNA encoding cancer associated antigen KOC-3.
XX
KW   Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.
XX
OS   Homo sapiens.
```

XX
PN WO9954738-A1.
XX
PD 28-OCT-1999.
XX
PF 16-MAR-1999; 99WO-US005766.
XX
PR 17-APR-1998; 98US-00061709.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR WPI; 2000-013284/01.
XX
PT Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers.
XX
PS Claim 55; Page 42-43; 44pp; English.
XX
CC The present sequence represents an alternative form of a cancer
CC associated antigen gene designated KOC-3. The specification also
CC describes a cancer associated antigen designated CT7. The CT7
CC polynucleotide was isolated from SK-MEL-37 melanoma cells. The
CC polypeptide has some homology with MAGE-10, limited to about 210 carboxy
CC terminal amino acids. The amino terminal of the protein has a repetitive
CC pattern, with repeats rich in serine, proline, glutamine and leucine, and
CC an almost invariable core of the peptide given in AAY43877. The CT7
CC polypeptide can be processed to peptides which provoke lysis by cytolytic
CC T cells. The polynucleotides and polypeptides can be used for treating a
CC cancerous condition and screening for or diagnosing cancerous conditions.
CC The cancer associated antigens can be used as an immunogenic or vaccine
CC composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte
CC macrophage-colony stimulating factor (GM-CSF)
XX
SQ Sequence 3283 BP; 945 A; 833 C; 779 G; 725 T; 0 U; 1 Other;

Query Match   92.1%; Score 3143; DB 3; Length 3283;
Best Local Similarity   96.2%;
Matches 3283; Conservative   0; Mismatches   0; Indels 129; Gaps   1;

Qy   1 GGCAGCGGAGGAGGCGGAGGAGGCGGGGAGGCCGGGGGGAGGCGGGGAGCCGGGGAGCCGGGG   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCAGCGGAGGAGGCGGAGGAGGCGGGGAGGCCGGGGGGAGGCGGGGAGCCGGGGAGCCGGGG   60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGACAACTCGGAGCCCCGTCACCCGCCGAC   120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGACAACTCGGAGCCCCGTCACCCGCCGAC   120

Qy   121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGCGGGACAGGTCCTGCTGAAG   180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGCGGGACAGGTCCTGCTGAAG   180

Qy   181 TCCGGGCTACGGCCTTCGTGGACTACCCCGGGACAGTGGGACAACTCGCGGCCATGGGCCATCGGGACAGC   240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGGCTACGGCCTTCGTGGACTACCCCGGGACAGTGGGACAACTCGCGGCCATGGGCCATCGGGACAGC   240

Qy   241 CTCTCGGGTAAAGTGAAGTGCATGGGAAAAATCATGGGAGAAGTTGATTGATTCGAAGTTCAGTCTCTAAA   300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGAAGTGCATGGGAAAAATCATGGGAGAAGTTGATTGATTCGAAGTTCAGTCTCTAAA   300

Qy   301 AAGCTAAGGAGCAGGAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCGAAGTGGGAGGTG   360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGCTAAGGAGCAGGAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCGAAGTGGGAGGTG   360

Qy   361 TTGGATGGACTTTGGCTCAATATGGGACCAGTGGGACAAGTCAACACAGAC   420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTGGCTCAATATGGGACCAGTGGGACAAGTCAACACAGAC   420

Qy   421 ACAGAAACCGCCGTTGTCAAGTCACATATGGCAAGAGAGAAGCAAAATAGCCATG   480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAAGTCACATATGGCAAGAGAGAAGCAAAATAGCCATG   480

Qy   481 GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCCTTCTACATCCGGAT   540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCCTTCTACATCCGGAT   540

Qy   541 GAAGAGGTGAGCTCCCCTTCGCCCCTGCCCCCTCAGCGACCCCAGCGTGGGGACCACTCTTCCCGG   600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCGCCCCTGCCCCCTCAGCGACCCCAGCGTGGGGACCACTCTTCCCGG   600

Qy   601 GAGCAAGAGGCCAGCGCCCTGGGGGCACTTCTCAGCGCCAGACAGATTGATTCCCGCTGCGG   660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGAGGCCAGCGCCCTGGGGGCACTTCTCAGCGCCAGACAGATTGATTCCCGCTGCGG   660

Qy   661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAAGAGGGGCTTGACCATAAAG   720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAAGAGGGGCTTGACCATAAAG   720

Qy   721 AACATCACTAAGCACAGAACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAAACTCTGGAGCT   780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCACAGAACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAAACTCTGGAGCT   780

Qy   781 GCAGAGAAGCTGTCACCATCCATGTCCACCCCAGAGGGACTTCTGAGCATGCGCATG   840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCTGTCACCATCCATGTCCACCCCAGAGGGGACTTCTGAAGCATGCGCATG   840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACGCAAACTAGCCAGGAGATTCCTCTG   900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGCAGATGAGACGCAAACTAGCCAGGAGATTCCTCTG   900

Qy   901 AAAATCTTGCCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAAGAAGGCAGAAATTTG   960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGCCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAAGAAGGCAGAAATTTG   960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGAATAACAAATCTCATCTTTGCAGGATTTGAGC   1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGAATAACAAATCTCATCTTTGCAGGATTTGAGC   1020

Qy   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGCCACAGTTGAGGCCTGTGCCAGTGCT   1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATATACAACCCGGAAAGAACCATCACTGTGAAGGCCACAGTTGAGGCCTGTGCCAGTGCT   1080

Qy   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCGGCTGTGTTAAC   1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAGATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAATGATATGCGGCTGTGTTAAC   1140

Qy   1141 CAACAAGCCAATCTGATCCGACGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA   1200
Db   1141 --   1140

Qy   1201 CTGTCCGTCGTGTATCTCCACCCAGCAGGGGCTCCCCCGTCCCCCGTCGCCCCCTACCAC   1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGTCCGTCGTGTATCTCCACCCAGCAGGGGCTCCCCCGTCCCCCGTCGCCCCCTACCAC   1260

Qy   1141 --   1140

Qy   1261 CCCTTCACTACCCACTCCGATACTCTCCGACTTGCCCCATCACCAGTTTGGCCCG   1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ----ACCCACTCCGATACTCTTCCGACTTGCCGTGTACCCCATCACCAGTTTGGCCCG   1191

Qy   1321 TTCCGCATCATCATCTATCCGAGAGGAGATTGTGAATCTCTTCATCCAACCCCAG   1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1192 TTCCGCATCATCATCTTATCCGAGAGCAGCAGAGAGATTGTGAATCTTTCATCCAACCCCAG   1251

Qy   1381 GCTGTGGGGCCGCAGCCTTTGGGGACGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCC   1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1252 GCTGTGGGGCCGCAGCCTTTGGGGACGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCC   1311

Qy   1441 GGAGCCTCTATCAAGATTGCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC   1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1312 GGAGCCTCTATCAAGATTGCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC   1371

Qy   1501 ATCACGGGCCACCGGAAGCCAGTTCAGGGCCACGGACGATCTTTGGGAAACTGAAA   1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1372 ATCACGGGCCACCGGAAGCCAGTTCAGGGCCACGGACGATCTTTGGGAAACTGAAA   1431

Qy   1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAGAGTGAGAGCTGAAGCCGCATATCGAGTGCCC   1620

```
Db   ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
     GAGGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGAAGCGCCATATCAGACGTGCCC 1491

Qy   TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1680
     ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1551

Qy   TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCAGATGAAAATGAGAGAAGTG 1740
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCAGATGAAAATGAGAGAAGTG 1611

Qy   ATCGTCAGAATTATCGGGCACTTCTTGCTAGCAGGACTGCACAGCGCAAGATCAGGGAA 1800
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ATCGTCAGAATTATCGGGCACTTCTTGCTAGCAGGACTGCACAGCGCAAGATCAGGGA 1671

Qy   ATTGTACAACAGGTGAGCAGCAGCAGAAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ATTGTACAACAGGTGAGCAGCAGCAGAAAATACCCTCAGGGAGTCGCCTCACAGCGC 1731

Qy   AGCAAGTGAGGCTCCCACAGCACCACAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   AGCAAGTGAGGCTCCCACAGCACCACAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1791

Qy   CTGACAGAATGAGACCAAACGCAGCCAGCGATCGGGAGCAAACCAAAGACCATCTGAG 1980
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CTGACAGAATGAGACCAAACGCAGCCAGCGATCGGGAGCAAACCAAAGACCATCTGAG 1851

Qy   GAATGAGAAGCTCTGCGGAGGGGCGCCAGGAGTCTGCGCAGGACTCTGCGAGACCCCCAGGGGC 2040
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GAATGAGAAGCTCTGCGGAGGGGCGCCAGGAGTCTGCGCAGGACTCTGCGAGACCCCCAGGGGC 1911

Qy   CGAGGAGGGGGCGCGGGAAGGTTCAGCCAGGTTGCCAGAGAACCACCACCGAGCCCTCCCGCC 2100
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CGAGGAGGGGGCGCGGGAAGGTTCAGCCAGGTTGCCAGAGAACCACCACCGAGCCCTCCCGCC 1971

Qy   CCCCAGGGGCCTTCTGCGAGGCTTCAGCGCATCCACCACTTCACCATCCTCGGATCTCTCCTGAA 2160
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CCCCAGGGGCCTTCTGCGAGGCTTCAGCGCATCCACCACTTCACCATCCTCGGATCTCTCCTGAA 2031

Qy   CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2091

Qy   AAAATGCACACCCTTTTCTGTGGCAAATCGTCCTGTACATGTGTACATATTAGAAA 2280
     |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   AAAATGCACACCCTTTTCTGTGGCAAATCGTCCTGTACATGTGTACATATTAGAAA 2151

Qy   GGGGAAGATGTTAAGATGTGCCTGTGGGTTACACAGGGTGCCTGCACGCGGTAATATAT 2340
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GGGGAAGATGTTAAGATGTGCCTGTGGGTTACACAGGGTGCCTGCACGCGGTAATATAT 2211

Qy   TTTAGAAATAATATCAAATAAATAACTCAACTAACTCCAATTTTAATCAATTATAAATTTT 2400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   TTTAGAAATAATATCAAATAAATAACTCAACTAACTCCAATTTTAATCAATTATAAATTTT 2271

Qy   TTTTTCTTTTTAAAGAAGAAGCACAGCTTTCTTAGAACTTTAAAGAAATAAAGTCTTTGGGAG 2460
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   TTTTTCTTTTTAAAGAAGAAGCACAGCTTTCTTAGAACTTTAAAGAAATAAAGTCTTTGGGAG 2331

Qy   GTCTCACGGTGTAGAGAGGAGCTTTGAGGCACCACCGCACAAAATTCACCCAGAGGGAAAT 2520
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GTCTCACGGTGTAGAGAGGAGCTTTGAGGCACCACCGCACAAAATTCACCCAGAGGGAAAT 2391

Qy   CTCGTCGGAAGGACACTCACGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA 2580
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CTCGTCGGAAGGACACTCACGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA 2451

Qy   CCGTCTCCTTGAAGAGGAAACTCTGTCACTCGTCTAGCTGTCACACCATT 2640
     |||||||||||||||||||||||||||||||||||||||||||||||||
Db   CCGTCTCCTTGAAGAGGAAACTCTGTCACTCGTCTAGCTGTCACACCATT 2511

Qy   TCTCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATATCGTCTATATAATTCTCTGTCT 2700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   TCTCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCT 2571

Qy   CTCTCTGTTTATCTCCCCTCCCCCTTCTTCCTCCATTCTCTTTGAA 2760
     |||||||||||||||||||||||||||||||||||| |||||||||
Db   CTCTCTGTTTATCTCCCCTCCCCCTTCTTCCTCCATTCTCTTTGAA 2631

Qy   TTTCCTCATCCCTTCCATCTCAATCCCGTATCATCCCCCCCCCCCCCAGGCAAAGCA 2820
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   TTTCCTCATCCCTTCCATCTCAATCCCGTATCATCACGCACCCCCCCCCCAGGCAAAGCA 2691

Qy   GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCAGCCTCAA 2880
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCGAAACACACAGCAGCCTCAA 2751

Qy   CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2940
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CTTACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAG 2811

Qy   AGGAAAACAGGAACCCACCAACCAAACCAAACAAATCAAGAAAAAATTCCACAATG 3000
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   AGGAAAACAGGAACCCACCAACCAAACCAAACAAATCAAGAAAAAATTCCACAATG 2871

Qy   AAAGAAATGTATTTTGTCTTTTGTGTATAAGCCATCAATATTCAGCAAAATG 3060
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   AAAGAAATGTATTTTGTCTTTTGTGTATAAGCCATCAATATTCAGCAAAATG 2931

Qy   ATTCCTTTCTTTAAAAAAAAAATGTGGAGAAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ATTCCTTTCTTTAAAAAAAAAATGTGGAGAAAAGTAGAAATTTACCAAGGTTGTTGGCC 2991

Qy   CAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAACACAGAGAAGCTACTTCAG 3180
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAACACAGAGAAGCTACTTCAG 3051

Qy   GTGTTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3240
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GTGTTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG 3111

Qy   TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3300
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACT 3171

Qy   GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACGTGTTCTGAATTGTCTTCCGCT 3360
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACGTGTTCTGAATTGTCTTCCGCT 3231

Qy   AGCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3412
     ||||||||||||||||||||||||||||||||||||||||||||||
Db   AGCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

```
RESULT 5
AAS70982
ID   AAS70982  standard; cDNA; 3667 BP.
XX
AC   AAS70982;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6786.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
```

PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06795.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 6786; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;

Query Match   91.3%; Score 3114.8; DB 5; Length 3667;
Best Local Similarity   96.1%; Pred. No. 0;
Matches 3277;   Conservative   0; Mismatches   3;   Indels 131; Gaps   3;

Qy   3 CAGCGGAGGAGGCGAGGAGGCCCGGGTACCGGCCGGGGAGCCCGGGGCCTCCGGGGAA   62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   369 CAACGGAGGAGGCGAGGAGGCCCGGGTACCGGCCGGGGAGCCCGGGGCCTCCGGGGAA 428

Qy   63 GAGACCGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCCGTCACCGCCGACGA   122
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   429 GAGACCGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCCGTCACCGCCGACGA 488

Qy   123 CCTCCGCAGCTCTTTGGGACAGGAGAAGTCGCCCCTGGGGGACAGGTCCTGCTGAAGTC 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CCTCCGCAGCTCTTTGGGACAGGAGAAGTCGCCCCTGGGGGACAGGTCCTGCTGAAGTC 548

Qy   183 CGGCTACGCCTTCGTGACTACCCCGACCAGAACTGGCATGGCCATCCGCGCCATCGAGACCCT 242
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   549 CGGCTACGCCTTCGTGACTACCCCGACCAGAACTGGCATGGCCATCGGCGCCATCGAGACCCT 608

Qy   243 CTCGGGTAAAGTGGAATTGGGGACAGTTGATTATTGGGAAAATCATGAAGTTGATTACTCAGTTCTAAAAA 302
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   609 CTCGGGTAAAGTGGAATTGCATGGGGAAAATCATGAAGTTGATTACTCAGTTCTAAAAA 668

Qy   303 GCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCACTTGCAGTTGGGAGGTGTT 362
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669 GCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCACTTGCAGTTGGGAGGTGTT 728

Qy   363 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACAC 422
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   729 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACAC 788

Qy   423 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAAATAGCCATGGA 482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   789 AGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAGAAGAAAATAGCCATGGA 848

Qy   483 GAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTCAAGATTC 542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   849 GAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTCACTACTCCGGATGA 908

Qy   543 AGAGGTGAGCTCCCCTTCGCCCTCAGCGAGCCCAGGCGGGGGGACCCACTCTCCGGGA 602
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909 AGAGGTGAGTCCCTTCGCCCTCAGCGAGCCCAGGCGGGGGGACCCACTCTCCGGGA 968

Qy   603 GCAAGGCCACGCCCGGGGCACTTCTCAGGCAGCAGATTGATTCCGCTGCGGAT 662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   969 GCAAGGCCACGCCCGGGGCACTTCTCAGGCAGCAGATTGATTCCCGCTGCGGAT 1028

Qy   663 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAA 722
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAA 1088

Qy   723 CATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAGAAGAACTCTGGAGCTGC 782
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAGAAGAACTCTGGAGCTGC 1148

Qy   783 AGAGAAGCCTGTCACCATCATGCACCCCAGAGAGGACTTCTGAAGCATGCCGCATGAT 842
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCCTGTCACCATCATGCACCCCAGAGAGGACTTCTGAAGCATGCCGCATGAT 1208

Qy   843 TCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 902
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy   903 AATCTTGGCACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTGAA 962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGGCAGAAATTTGAA 1328

Qy   963 GAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1022
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1388

Qy   1023 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGTGCCAGTGCTGA 1082
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGCCTGTGCCAGTGCTGA 1448

Qy   1083 GATAGAGATTATGAAGAAGCTGCGTGAGCCCTTGAAAATGATATGCTGCTGTTAACCA 1142
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1449 GATAGAGATTATGAAGAAGCTGCGTGAGCCCTTTGAAAATGATATGCTGCTGTTAAC-- 1506

Qy   1143 ACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACT 1202
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 -- 1506

Qy   1203 GTCCGTGCTATCTCCACCAGCAGGCCCCGCCCTCCCCCGCTGCCCCCTACCACCC 1262

Db   1507 -- 1506

Qy   1263 CTTCACTACCCACTCCGGATACTTCTCCAGCAGTTGGCCCATCACCAGTTGGCCCGTT 1322
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ------ACCCACTCCGGATACTTCTCCAGCAGTTGGCCCATCACCAGTTGGCCCGTT 1559

Qy   1323 CCCGCATCATCACTCTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGC 1382
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGC 1619

Qy   1383 TGTGGGGCGCCATCATCGGGAAGAGGGGGGCACACATCAAACAGTGGCGAGATTCGCCGG 1442
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 TGTGGGGCGCCATCATCGGGAAGAGGGGGGCACACATCAAACAGTGGCGAGATTCGCCGG 1679

Qy   1443 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGAGTCAGCGAAGGATGGTCATCAT 1502
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 AGCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGAGTCAGCGAAGGATGGTCATCAT 1739

Qy   1503 CACCGGCGCACCGAAGCCCAGTTCAAGGCCACAGTTCAAGGCGACCGGATCTTTGGGAAACTGAAAGA 1562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   1740 CACCGGGCCACCGGAGGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGA 1799
Qy   1563 GGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCGAGTGCCCTC 1622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 GGAAAACTTCTTTAACCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCGAGTGCCCTC 1859
Qy   1623 TTCCACAGTCTGCCGGGTGATTGGCAAGACCGTGACGAACTGCGACAACTT 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 TTCCACAGTCTGCCGGGTGATTGGCAAGACCGTGACGAACTGCGACAACTT 1919
Qy   1683 AACCAGTGCAGAAGTCATCGTGCCTCGTGACAAAACGCCAGATGAAAATGAGGAAGTGAT 1742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 AACCAGTGCAGAAGTCATCGTGCCTCGTGACAAAACGCCAGATGAAAATGAGGAAGTGAT 1979
Qy   1743 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGGGCAAGATCAGGGAAAT 1802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGGGCAAGATCAGGGAAAT 2039
Qy   1803 TGTACAACAGTGAAGCAGCAGCAGAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG 1862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 TGTACAACAGTGAAGCAGCAGCAGAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG 2099
Qy   1863 CAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCT 1922
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 CAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGATGAATGTAGCCCTTCCAACACCT 2159
Qy   1923 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACAAAGACCATCTGAGA 1982
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACAAAGACCATCTGAGA 2219
Qy   1983 ATGAGAAGTCTGCGGAGGCGGCGCCAGGGACTCGGAGGACTCTGCCGAGCCCTGAGAACC 2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220 ATGAGAAGTCTGCGGAGGCGGCGCCAGGGACTCGGAGGACTCTGCCGAGCCCTGAGAACC 2279
Qy   2043 AGGGGCGGGGGGAAGTCAGCGGCTTCAGCCATCCACTCACTTGTTTGCCGA 2102
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280 AGGGGCGGGGGGAAGTCAGCGGCTTCAGCCATCCACTCACTTGTTTGCCGA 2339
Qy   2103 CCAGGGCTTCTGCAGCTTCAGCCTTTTAGTTGAACTAACATAGTGAACT 2162
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2340 CCAGGGCTTCTGCAGCTTCAGCCTTTTAGTTGAACTAACATAGTGAACT 2399
Qy   2163 CCCAGACGCTATCCCTTTAGTTGAACTAACATAGTGAAGCTGTTCAAAGCCAAGCAA 2222
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400 CCCACACGCTATCCCTTTAGTTGAACTAACATAGTGAAGCTGTTCAAAGCCAAGCAA 2459
Qy   2223 AATGCACACCTTTTTCTGTGGCAAATCGTCTCTGTACACATGTGTACATATTAGAAAGG 2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2460 AATGCACACCTTTTTCTGTGGCAAATCGTCTCTGTACACATGTGTACATATTAGAAAGG 2519
Qy   2283 GAAGATGTTAAGATATGTGGCCTGGGGTTACACAGGGTGCCTGCAGGGGTAATATATTT 2342
          |||||||||||||||||||||||||||||||||||||||||||
Db   2520 GAAGATGTTAAGATATGTGGCCTGGGGTTACACAGGGTGCCTGCAGGGGTAATATATTT 2579
Qy   2343 TAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTT 2402
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2580 TAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTTTT 2639
Qy   2403 TTTCTTTTTAAAGAAGAAAGCAGAGGCTTTGAGGCTTTCTGAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2462
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2640 TTTCTTTTTAAAGAAGAAAGCAGAGGCTTTGAGGCTTTCTGAGACTTTAAAGAATAAAGTCTTTGGGAGGT 2699
Qy   2463 CTCACGGTGTGAGAGGAGACTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCT 2522
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2700 CTCACGGTGTGAGAGGAGACTTTGAGGCCACCCGCACAAATTCACCCAGAGGGAAATCT 2759
Qy   2523 CGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTATGTCAACAGAAGGGGATACC 2582
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2760 CGTCGGAAGGACACTCACGGCAGTTCTGGATCACGTGTATGTCAACAGAAGGGGATACC 2819
Qy   2583 GTCTCCTTGAAGAGGAAACTTCATTCACTCCTCAGCTCATGCCTGCTCAGCTCATACACCCCATTTC 2642
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2820 GTCTCCTTGAAGAGGAAACTTCATTCACTCCTCAGCTCATGCCTGCTCAGCTCATACACCCCATTTC 2879
```

```
Db   2643 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2880 TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT 2939
Db   2703 CTCTGTTTATCTCTCCCTCCCCCTCCCTTTCTCTCCATCTCCATTCTTTTTGAATT 2762
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2940 CTCTGTTTATCTCTCCCTCCCCCTCCCTTTCTCTCCATCTCCATTCTTTTTGAATT 2999
Db   2763 TCCTCATCCTCCTCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCAGGCAAAGCAGT 2822
          ||||||||||||||||||||||||||||||||||||||||                 ||||
Qy   3000 TCCTCATCCTCCTCCATCTCAATCCCGTATCTACGGCA-CCCCCCCCCCAGGCAAAGCAGT 3058
Db   2823 GCTCTGAGTATCATCACACAAAAGGAACAAAAGGAACAAAAGGAAACACACAAACCAGCCTCAACT 2882
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3059 GCTCTGAGTATCACATCACACAAAAGGAACAAAAGGAAACACACAAACCAGCCTCAACT 3118
Db   2883 TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTCGTGTCCTAGCGTTTTGGAAGAG 2942
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3119 TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTCGTGTCCTAGCGTTTTGGAAGAG 3178
Db   2943 GAAAACACGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAA 3002
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3179 GAAAACACGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAA 3238
Db   3003 AGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3062
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3239 AGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT 3298
Db   3063 TCCTTTCTTT-AAAAAAAAAATGTGAGGAGAAGTAGAAATTACCAAGGTGTTGGCCC 3121
          ||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Qy   3299 TCCTTTCTTTAAAAAAAAAAATGTGAGGAGAAGTAGAAATTACCAAGGTGTTGGCCC 3358
Db   3122 AGGGCGTAAATCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGG 3181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3359 AGGGCGTAAATCACAGATTTTTTAACGAGAAAAACACACAGAGAAGAAGCTACCTCAGG 3418
Db   3182 TGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGT 3241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3419 TGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAGT 3478
Db   3242 TGGAGCATTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTG 3301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3479 TGTTTTTTACCTCAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTG 3538
Db   3302 CCAGCCTGAGAGGTGACAGTCCAAGTGTCAACAGCGTGTTCGAATTGTCTTCCGCTA 3361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3539 CCAGCCTGAGAGGTGACAGTCCAAGTGTCAACAGCGTGTTCGAATTGTCTTCCGCTA 3598
Db   3362 GCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          ||||||||| |||||||||||||||||||||||||||||||||||||||||
Qy   3599 GCCAAGAACCTATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT 3649
```

RESULT 6
ABS76442
ID   ABS76442 standard; cDNA; 3667 BP.
XX
AC   ABS76442;
XX
DT   11-DEC-2002 (first entry)
XX
DE   cDNA encoding human ovarian cancer marker M452.
XX
KW   Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW   central nervous system disorder; bacterial meningitis; viral meningitis;
KW   Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW   brain herniation; inflammation; encephalitis; testicular disorder;
KW   nontuberculous granulomatous orchitis; connective tissue disorder;
KW   heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW   histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX
OS   Homo sapiens.

```
XX
PN   WO200271928-A2.
XX
PD   19-SEP-2002.
XX
PF   14-MAR-2002; 2002WO-US007826.
XX
PR   14-MAR-2001; 2001US-0276025P.
PR   14-MAR-2001; 2001US-0276026P.
PR   10-AUG-2001; 2001US-0311732P.
PR   19-SEP-2001; 2001US-0323580P.
PR   26-SEP-2001; 2001US-0324967P.
PR   26-SEP-2001; 2001US-0325102P.
PR   26-SEP-2001; 2001US-0325149P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI   Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
PI   Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
DR   WPI: 2002-723277/78.
DR   P-PSDB: ABG96346.
XX
PT   Assessing whether a patient is afflicted with ovarian cancer, useful in
PT   assessing the stage or progression of the disease, comprises comparing
PT   the expression level of a cancer marker in a sample from a patient and
PT   from a non cancer patient.
XX
PS   Disclosure; Page 262-263; 481pp; English.
XX
CC   The present invention relates to a new method for assessing whether a
CC   patient is afflicted with ovarian cancer. The method involves comparing
CC   the expression level of a marker in a patient sample and the normal level
CC   of expression of the marker in a control non-ovarian cancer sample, where
CC   the marker is selected from 363 cancer markers described in the
CC   specification. The method of the invention is useful in diagnosing or
CC   characterising cancer, in detecting the presence of cancer as early as
CC   possible, and the recurrence of ovarian cancer. The method may also be of
CC   particular use with patients having an enhanced risk of developing
CC   ovarian cancer (e.g. patients having a familial history of ovarian
CC   cancer). The cancer markers may be used in the management and treatment
CC   of e.g. brain and central nervous system disorders (e.g. bacterial and
CC   viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC   disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
CC   inflamations (e.g. bacterial or viral meningitis or encephalitis),
CC   testicular disorders (e.g. nontuberculous granulomatous orchitis),
CC   connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC   disease or atherosclerosis). The compositions and methods may also be
CC   used in assessing the histological type of neoplasm associated with
CC   ovarian cancer, monitoring the progression of ovarian cancer, determining
CC   whether ovarian cancer has metastasized or is likely to metastasize,
CC   selecting a composition for inhibiting ovarian cancer, assessing the
CC   ovarian carcinogenic potential of a compound, or inhibiting ovarian
CC   cancer or at risk of developing ovarian cancer. The present nucleic acid
CC   sequence encodes one of the ovarian cancer markers described in the
CC   invention
XX
SQ   Sequence 3667 BP; 1081 A; 882 C; 846 G; 858 T; 0 U; 0 Other;

Query Match      91.3%;  Score 3114.8;  DB 6;  Length 3667;
Best Local Similarity  96.1%;  Pred. No. 0;
Matches 3277;  Conservative  0;  Mismatches  3;  Indels 131;  Gaps  3;
```

```
Qy      3 CAGCGGCGAGGAGGCGAGGAGCGCCGGGTACCCGCCGCGGGGAGCCCGGGCTCTCGGGGAA  62
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    369 CAACGGCGAGGAGGCGAGGAGCGCCGGGTACCCGCCGCGGGGAGCCCGGGCTCTCGGGGAA 428

Qy     63 GAGACCGGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCCGACGA             122
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GAGACCGGATGATGAACAAGCTTTACATCGGGAACTTGAGCCCGTCACCGGCCGCCGACGA 488

Qy    123 CCTCCCGGCCAGCTCTTTGGGGACAGGAAGCTGCCCCTCGGCGGGACAGGTCTCTGCTGAAGTC 182
```

```
Db    489 CCTCCGGCCAGCTCTTTGGGGACACGGAACTGCCCCTGGGCGGGGCCATCGGCCATCGAGACCCT  548
          ||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||||
Qy    183 CGGCTACGGCCTTCGTTGGGACTACCCGACCAGAACTGGGCCATCGGCCATCCGCGCCATCGAGACCCT  242

Db    549 CGGCTACGGCCTTCGTTGGGACTACCCGACCAGAACTGGGCCATCGGCCATCCGCGCCATCGAGACCCT  608
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    243 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA  302

Db    609 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA  668
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    303 GCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGGAGGTGTT  362

Db    669 GCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGGAGGTGTT  728

Qy    363 GGATGGACTTTGGCTCAATATGGACACGTGGACAATGTGGAACAAGTCAACACAGACAC  422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GGATGGACTTTGGCTCAATATGGACACGTGGACAGTGGACAATGTGGAACAAGTCAACACAAGACAC  788

Qy    423 AGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGGAGCAAAAATAGCCATGGA  482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAGGAGCAAAAATAGCCATGGA  848

Qy    483 GAAGTAAGCGGGCATCAGTTTGAGAACTTCCTTCAAGATTCTTCAAGATTCCTACATCCCGGATGA  542
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GAAGTAAGCGGGCATCAGTTTGAGAACTTCCTTCAAGATTCCTACATCCCGGATGA  908

Qy    543 AGAGGTGAGTCCCCTTCGCCCCTCGCCCCCCTCAGCGGAGCCCAGCGTGGGGACCACTCTTCCCGGGA  602
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 AGAGGTGAGTCCCCTTCGCCCCTCGCCCCCCTCAGCGGAGCCCAGCGTGGGGACCACTCTTCCCGGGA  968

Qy    603 GCAAGGCCACGCCCTGGGGCCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGAT  662
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GCAAGGCCACGCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGAT 1028

Qy    663 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAA  722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CCTGGTCCCCACCAGTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAA 1088

Qy    723 CATCACTAAGCAGACAGCCAGTCCCGGTAGATATCCATAGAGAAAGAGAACTCTGGAGCTGC  782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CATCACTAAGCAGACAGCCAGTCCCGGTAGATATCCATAGAGAAAGAGAACTCTGGAGCTGC 1148

Qy    783 AGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCCATGAT  842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 AGAGAAGCCTGTCACCATCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCCATGAT 1208

Qy    843 TCTTGAAATCATGCGAAGAGGCAGATGAGACCAAAACTAGCCGAAGAGATTCCTCTGAA  902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TCTTGAAATCATGCGAAGAGGCAGATGAGACCAAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy    903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTGAA  962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAAGGCAGAAATTTGAA 1328

Qy    963 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1388

Qy   1023 ATACAACCCGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGA 1082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 ATACAACCCGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCTGA 1448

Qy   1083 GATACAGGATTATGAGAAGAACTGCTTGAGGCCTTTGAAAATGATATGCTGCTGTTAACCA 1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   1449 GATAGAGATTTATGAGAAGAAGTGCGTGAGGCCTTTGAAAATGATATGCTGCTGTTAAC-- 1506

Qy   1143 ACAAGCCAATCTGATCCCAGGGTTGAACCTTGGCATCTTTTCAAGAGGACT 1202
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1507 ----------------------------------------------------- 1506

Qy   1203 GTCCGTGCTATCTCCACCAGCAGGGCCCCCGGGAGCTCCCCCGCTGCCCCTACCACCC 1262
```

Right-hand alignment block:

```
Db  1507  --------------------------------------------------------------  1506

Qy  2402  TAGAAATATATATCAAATAACTAACTCCAATTTTTAATCAATTATTAATTTTT  2343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2639  TAGAAATATATATCAAATAACTAACTCCAATTTTTAATCAATTATTAATTTTT  2580

Qy  2462  TTTCTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT  2403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2699  TTTCTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT  2640

Qy  2522  CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCAGAGGGAAATCT  2463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2759  CTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCAGAGGGAAATCT  2700

Qy  2582  CGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTATGTCAACAGAGGGATACC  2523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2819  CGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTATGTCAACAGAGGGATACC  2760

Qy  2642  GTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCTAGCTCATCACCCATTTC  2583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2879  GTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCTAGCTCATCACCCATTTC  2820

Qy  2702  TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT  2643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2939  TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT  2880

Qy  2762  CTCTGTTTATCTCTCCCCTCCCCCCCCCCCCCCCCCCATCTCCATTCTCTTTGAATT  2703
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  2999  CTCTGTTTATCTCTCCCCTCCCCCCCCCCCCCTTCTCCATTCTCTTTGAATT  2940

Qy  2822  TCCTCATCCCTCCATCTCAATCCGTATCTACGCACCCCCCCCCCCCATCTCCAGGCAAAGCAGT  2763
          |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db  3058  TCCTCATCCCTCCATCTCAATCCGTATCTACGCA-CCCCCCCCCCCCAGGCAAAGCAGT  3000

Qy  2882  GCTCTGAGTATCACATCACACAAAAGCCAAACACACAACCAGCCTCAACT  2823
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3118  GCTCTGAGTATCACATCACACAAAAGCGAAACACACAACCAGCCTCAACT  3059

Qy  2942  TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG  2883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3178  TACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG  3119

Qy  3002  GAAAACAGGAACCCACCAAACCAACCAATCAACAACAAACAAAGAAAAAATTCCACAATGAA  2943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3238  GAAAACAGGAACCCACCAAACCAACCAATCAACAACAAACAAAGAAAAAATTCCACAATGAA  3179

Qy  3062  AGAATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT  3003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3298  AGAATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT  3239

Qy  3121  TCCTTTCTTT-AAAAAAAAAATGTGGAAGAAAGTAGAAATTTACCAAGGTTGTTCGCCC  3063
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  3358  TCCTTTCTTTAAAAAAAAAAATGTGGAAGAAAGTAGAAATTTACCAAGGTTGTTCGCCC  3299

Qy  3181  AGGGCGTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAAGTCACTTCAGG  3122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3418  AGGGCGTAAATTCACAGATTTTTTTAACGAGAAAAACACACAGAGAAGTCACTTCAGG  3359

Qy  3241  TGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTGAGATTTTCCCTTAGAGCTGATAGT  3182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3478  TGTTTTTACCTCAGCACCTTGCTCTTGTGTTTCCCTTGAGATTTTCCCTTAGAGCTGATAGT  3419

Qy  3301  TGGAGCATTTTTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3242
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db  3538  TGGAGCATTTT---TTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACTG  3479

Qy  3361  CCAAGAACCAAGTTGACAGTCCAAGTGTCGCAACAGCTGTTCTGAATTGTCTTCCGCTA  3302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3598  CCAAGAACCAAGTTGACAGTCCAAGTGTCGCAACAGCTGTTCTGAATTGTCTTCCGCTA  3539

Qy  3412  GCCAAGAACCNATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT  3362
          |||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  3649  GCCAAGAACCTATATGGCCTTCTTCTTTGGACAAACCTTGAAAATGTTTATTT  3599
```

Left-hand alignment block:

```
Db  1507  --------------------------------------------------------------

Qy  1263  CTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTT  1322
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1507  ------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTGGCCCGTT  1559

Qy  1323  CCCGCATCACTCATCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCTCCAACCCAGGC  1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1560  CCCGCATCACTCATCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCTCCAACCCAGGC  1619

Qy  1383  TGTGGGCGCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGGCCGAGATTCGCCG  1442
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1620  TGTGGGCGCCATCATCGGGAAGAAGAGGGGGCACACATCAAACAGCTGGCCGAGATTCGCCG  1679

Qy  1443  AGCCCTATCAAGATTGCCCCTGCCGGAAGGCCCCAGACGTCAGCGAAAGGATGGTCATCAT  1502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1680  AGCCCTATCAAGATTGCCCCTGCCGGAAGGCCCCAGACGTCAGCGAAAGGATGGTCATCAT  1739

Qy  1503  CACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAAGA  1562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1740  CACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGGATCTTTGGGAAACTGAAAGA  1799

Qy  1563  GGAAAACTTCTTTAACCCAAAGAGAAGAAGTGAAGCGCATATCAGAGTGCCCTC  1622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1800  GGAAAACTTCTTTAACCCAAAGAGAAGAAGTGAAGCGCATATCAGAGTGCCCTC  1859

Qy  1623  TTCCACAGTCTGGCCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCGAGAACTT  1682
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1860  TTCCACAGTCTGGCCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCGAGAACTT  1919

Qy  1683  AACCAGTGCAGAAGTCATCGTGCCCTCGTGACCAAAGCGCCAGATGAAAATGAGGAAGTGAT  1742
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1920  AACCAGTGCAGAAGTCATCGTGCCCTCGTGACCAAAGCGCCAGATGAAAATGAGGAAGTGAT  1979

Qy  1743  CGTCAGAAATTATCGGGCACTTCTTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT  1802
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1980  CGTCAGAAATTATCGGGCACTTCTTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT  2039

Qy  1803  TGTACAACAGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  1862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2040  TGTACAACAGTGAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  2099

Qy  1863  CAAGTGAGGGTCCCACAGAGGCGGCCACAGGACCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACACCT  1922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2100  CAAGTGAGGGTCCCACAGAGGCGGCCACAGGACCACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACACCT  2159

Qy  1923  GACAGAGAATGAGAACCAAACCCAGCCAGCCAGCAGATCGGAGGCAAACCAAAGACCATCTGAGGA  1982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2160  GACAGAGAATGAGAACCAAACCCAGCCAGCCAGCCAGATCGGAGGCAAACCAAAGACCATCTGAGGA  2219

Qy  1983  ATGAGGGGGCGGCGGGAAGGTCAGCCAGGTTTGCCAGAACCACCAGGCCCTGAGAACCCCAGGGGCG  2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2220  ATGAGGGGGCGGCGGGAAGGTCAGCCAGGTTTGCCAGAACCACCAGGCCCTGAGAACCCCAGGGGCCG  2279

Qy  2043  AGGAGGGGGCCGGGAGGGGGGAAGGTCAGCCAGGTTTGCCAGAACCCCAGGGCGCCCC  2102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2280  AGGAGGGGGCCGGGAGGGGGGAAGGTCAGCCAGGTTTGCCAGAACCCCGAGCCCCGCGCCCC  2339

Qy  2103  CCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTGGAATCTCTGGAACT  2162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2340  CCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCGGATCTCTGGATCTCTGGAACT  2399

Qy  2163  CCCACGACGCTATCCCTTTTAGTTGAACTAACAGGTGAACGTGTTCAAAGCCAAGCAA  2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2400  CCCACGACGCTATCCCTTTTAGTTGAACTAACAGGTGAACGTGTTCAAAGCCAAGCAA  2459

Qy  2223  AATGCACACACCCTTTCTGGGTTACACAGGGTGCCTGTACAGTGTTACACATATTAGAGAAGG  2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2460  AATGCACACACCCTTTCTGGGTTACACAGGGTGCCTGTACAGTGTTACACATATTAGAGAAGG  2519

Qy  2283  GAAGATGTTAAGATGTGTGGGTTACACAGGTGCCTGCCTGCAGGCGTAATATATTT  2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2520  GAAGATGTTAAGATGTGTGGGTTACACAGGTGCCTGCCTGCAGGCGTAATATATTT  2579
```

```
RESULT 7
AAS26153
ID   AAS26153 standard; cDNA; 2290 BP.
XX
AC   AAS26153;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 332.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216447P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227009P.
PR   23-AUG-2000; 2000US-0227182P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.

PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234937P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0258365P.
PR   29-SEP-2000; 2000US-0263227P.
PR   29-SEP-2000; 2000US-0263367P.
PR   29-SEP-2000; 2000US-0263368P.
PR   29-SEP-2000; 2000US-0263369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
```

```
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251556P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
XX   WPI; 2001-488783/53.
DR   P-PSDB; AAU16166.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 332; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
```

```
Query Match             58.3%;  Score 1989.8;  DB 4;  Length 2290;
Best Local Similarity   99.4%;  Pred. No. 0;
Matches 2028;  Conservative   0;  Mismatches   0;  Indels   8;  Gaps   3;

Qy   1376  CCCAGGCTGTGGCGCCATCATCGGGAGGAAGGGGGCACACATCAAACAGTCGGCCAGGAT  1435
                ||| ||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db     25  CCCACGCGTCCGGCGCCATCATCGGGAGGAAGGGGGCACACATCAAACAGTCGGCCAGGAT    84

Qy   1436  TCGCCGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGCGAAAAGGATGG  1495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85  TCGCCGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCAGCGAAAAGGATGG   144

Qy   1496  TCATCATCACCGGCCACCGGACGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAAC  1555
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    145  TCATCATCACCGGCCACCGGAGGCCCAGTTCAAGGCCCAGGGACGATCTTTGGGAAAC   204
```

```
Qy   1556  TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGAAGTGGAGGCGCATATCAGAG  1615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205  TGAAAGAGAGAAACTTCTTTAACCCCAAAGAAGAAGAAGTGGAGGCGCATATCAGAG   264

Qy   1616  TGCCCTCTTCCACAGCTGCCGCCGGGTGATTGGCAAGACCGTGAACGAACTGC  1675
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265  TGCCCTCTTCCACAGCTGCCGCCGGGTGATTGGCAAGACCGTGAACGAACTGC   324

Qy   1676  AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG  1735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325  AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG   384

Qy   1736  AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCA  1795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385  AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCA   444

Qy   1796  GGGAAATTGTACAACAGTGAGCAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCAC  1855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445  GGGAAATTGTACAACAGTGAGCAGCAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTCAC   504

Qy   1856  AGGCAGCAAGTGAGGCTCCCACAGGCCACCAGCAACACACCGATGAATGAGCCCTTCC  1915
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505  AGGCAGCAAGTGAGGCTCCCACAGGCCACCAGCAACACACCGATGAATGAGCCCTTCC   564

Qy   1916  AACACCTGACAGAATGAGACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCAT  1975
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565  AACACCTGACAGAATGAGACCAGCCAGCCAGATCGGGAGCAAACCAAAGACCAT   624

Qy   1976  CTGAGGAATGAGAAGTCTGCGAGAGGCGGCCAGGGACCTCGCCGAGGCCCTGAGAACCCCA  2035
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625  CTGAGGAATGAGAAGTCTGCGAGAGGCGGCCAGGGACCTCGCCGAGGCCCTGAGAACCCCA   684

Qy   2036  GGGGCCGGAGCGGGCGGGGAAGTCAGCCAGTTGCCAGAACCACCGAGCCCGCCTC  2095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685  GGGGCCGGAGCGGGCGGGGAAGTCAGCCAGTTGCCAGAACCACCGAGCCCGCCTC   744

Qy   2096  CCGCCCCCAGGGCTTCTGCGAGGCTTCAGCGAGCTTCAGCCATCCACTTCACCATCGGATCTCTC  2155
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745  CCGCCCCCAGGGCTTCTGCGAGGCTTCAGCGAGCTTCAGCCATCCACTTCACCATCGGATCTCTC   804

Qy   2156  CTGAACTTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC  2215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805  CTGAACTTCCCACGACGCTATCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGC   864

Qy   2216  CAAGCAAATGCACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTACATATT  2275
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865  CAAGCAAATGCACACCCTTTTCTGTGGCAAATGTCTCTGTACATGTGTACATATT   924

Qy   2276  AGAAAGGGAAGAATGTTAAGATATGTGGCCTCGGTTACACAGGGTGCCTGCAGCGGTAA  2335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925  AGAAAGGGAAGAATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAA   984

Qy   2336  TATATTTTAGAAATAATATCAAATAACTCAACTAACTCAATTTTAATCAATTATTA  2395
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985  TATATTTTAGAAATAATATCAAATAACTCAACTAACTCAATTTTAATCAATTATTA   1044

Qy   2396  ATTTTTTTTCTTTTTAAAGAAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTT  2455
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1045  ATTTTTTTTCTTTTTAAAGAAGAAGCAGGCTTTTCTAGACTTTAAAGACTAAAGTCTTT   1104

Qy   2456  GGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCCGAGAG  2515
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105  GGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACCCGAGAG   1164

Qy   2516  GAAATCTCGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTATGTCAACAGAAG  2575
           |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165  GAAATTCGTCGGAAGGACACTCACGGCAGTTCTTGGATCACCTGTATGTCAACAGAAG   1224

Qy   2576  GGATACCGTCTCCTTGAAGAGGAAACTCTGTCACTCGTCGTCTGTCCAGCTCATCAC  2635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225  GGATACCGTCTCCTTGAAGAGGAAACTCTGTCACTCGTCGTCTGTCCAGCTCATCAC   1284
```

```
Qy    2636  CCATTTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTC  2695
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1285  CCATTTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTC  1344

Qy    2696  TGTCTCCTCGTTATCTCCCCCCCTCCCCCCTTCTTCCCACCTCCCATCTCCATTCTT  2755
            |||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    1345  TGTCTCCTCTG--TTATCTCCCCCCCTCCCCCCTTCTTCCCACCTCCCATCTCCATTCTT  1403

Qy    2756  TTGAATTTCCTCATCCCTCCATCTCCATCTCAATCCCGTATCTACGCA--CCCCCCCCCCCAGG  2813
            |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    1404  TTGAATTTCCTCATCCCTCCATCTCCATCTCAATCCCGTATCTACGCACCACCCCCCCCCCCCAGG  1463

Qy    2814  CAAAGCAGTCGCTCTGAGTATCATCACATCACAACAAAAGGAACAAAAGCGAAACACACAAACCA  2873
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1464  CAAAGCAGTCGCTCTGAGTATCATCACATCACAACAAAAGGAACAAAAGCGAAACACACAAACCA  1523

Qy    2874  GCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTT  2933
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1524  GCCTCAACTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTT  1583

Qy    2934  TTGGAAGAGGAAAACAGGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTC  2993
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1584  TTGGAAGAGGAAAACAGGAACCCACCAACCAATCAACCAAACAAAGAAAAAATTC  1643

Qy    2994  CACAATGAAAGAAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAG  3053
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1644  CACAATGAAAGAAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAG  1703

Qy    3054  CAAAATGATTCCTTTCTTT-AAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGT  3112
            ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1704  CAAAATGATTCCTTTCTTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGT  1763

Qy    3113  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAAAACCACAGGAAGAAGC  3172
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1764  TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAGAAAAACCACAGGAAGAAGC  1823

Qy    3173  TACCTCAGGTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAA  3232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1824  TACCTCAGGTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTTGTAAA  1883

Qy    3233  GCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAATAAGA  3292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1884  GCTGATAGTTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAATAAGA  1943

Qy    3293  TATCAACTGCCAGCCTGGAGAAGGTGACAGTGCAAGCAGCTGTCTCGAATTGT  3352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1944  TATCAACTGCCAGCCTGGAGAAGGTGACAGTGCAAGCAGCTGTCTCGAATTGT  2003

Qy    3353  CTTCCCTAGCCAAGAACCNATATGGCCCTCTTTTGGACAACACCTTGAAAATGTTTATTT  3412
            |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    2004  CTCCCGCTAGCCAAGAACCTATATGGCCTTCTTTTGGACAACACCTTGAAAATGTTTATTT  2063
```

```
RESULT 8
ABX73494
ID   ABX73494 standard; DNA; 2290 BP.
XX
AC   ABX73494;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #322.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
```

```
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001;  2001US-00764864.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   28-JUN-2000;  2000US-0214886P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0225447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   22-AUG-2000;  2000US-0226868P.
PR   30-AUG-2000;  2000US-0288924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   08-SEP-2000;  2000US-0231413P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0234997P.
PR   27-SEP-2000;  2000US-0235834P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241809P.
PR   01-NOV-2000;  2000US-0244617P.
PR   17-NOV-2000;  2000US-0249299P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
XX
PA   (ROSE/)  ROSEN C A.
PA   (RUBE/)  RUBEN S M.
PA   (BARA/)  BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55234.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 332; 402pp; English.
XX
```

CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
XX human novel polynucleotides of the invention
SQ Sequence 2290 BP; 707 A; 561 C; 491 G; 531 T; 0 U; 0 Other;

Query Match   58.3%; Score 1989.8; DB 7; Length 2290;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2028; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

Qy 1376 CCCAGGCTGTGGGGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTCGGCGAGAT 1435
        |||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 CCCACGCGTCCGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTCGGCGAGAT 84

Qy 1436 TCGCCGGAGCCTCTATCAAGATTGCCCTGCGGAAGGGCCCAGAGCGTCAGCAGAAGGATGG 1495
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 TCGCCGGAGCCTCTATCAAGATTGCCCTGCGGAAGGGCCCAGAGCGTCAGCAGAAGGATGG 144

Qy 1496 TCATCATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAC 1555
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 TCATCATCACCGGGCCACCCGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAC 204

Qy 1556 TGAAAGAGGAGAAACTTCTTTAACCCAAAGAACAAGAAGTGAAGCTGGAAGCGCTATATCAGAG 1615
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 TGAAAGAGGAGAAACTTCTTTAACCCAAAGAACAAGAAGTGAAGCTGGAAGCGCCATATCAGAG 264

Qy 1616 TGCCCTTCCACAGCTGGCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGC 1675
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 TGCCCTTCCACAGCTGGCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGC 324

Qy 1676 AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG 1735
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 325 AGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG 384

Qy 1736 AAGTGCATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGGCGCAAGATCA 1795
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 385 AAGTGCATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGGCGCAAGATCA 444

Qy 1796 GGGAAATTGTACACACAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTGCCTCAC 1855
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 445 GGGAAATTGTACACACAGGTGAAGCAGCAGGAGCAGGAGAAATACCCTCAGGGAGTCGCCTCAC 504

Qy 1856 AGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGGATGAATGTAGCCCCTTCC 1915
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 505 AGCGCAGCAAGTGAGGCTCCCACAGGCACCAGCAAAACAACGGATGAATGTAGCCCCTTCC 564

Qy 1916 AACACCTGACAGAATGAGACAAAACGCAGCAGCCAGTCGGAGGCAAACCAAAGACCAT 1975
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 565 AACACCTGACAGAATGAGACAAAACGCAGCAGCCAGTCGGAGGCAAACCAAAGACCAT 624

Qy 1976 CTGAGGAATGAGAAGTCTGCGGAGGGCGGCCAGGGACTTGCCTGCCGAGGCCCTGAGAACCCCA 2035
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 625 CTGAGGAATGAGAAGTCTGCGGAGGGCGGCCAGGGACTTGCCTGCCGAGGCCCTGAGAACCCCA 684

Qy 2036 GGGGCCGAGGAGGGCGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCCGCCTC 2095
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 685 GGGGCCGAGGAGGGCGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCCGCCTC 744

Qy 2096 CCGCCCCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTC 2155
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 745 CCGCCCCCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCATCCACTCGGATCTC 804

Qy 2156 CTGAACTCCACGACGCTATCCCTTTTAGTTGAACTAACTAACATAGGTGAACGTGTCAAAGC 2215
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 805 CTGAACTCCACGACGCTATCCCTTTTAGTTGAACTAACTAACATAGGTGAACGTGTCAAAGC 864

Qy 2216 CAAGCAAATGCACACCCTTTTCTGTGGCAAATGCTCTGTACATGTGTGTACATATT 2275
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 865 CAAGCAAATGCACACCCTTTTCTGTGGCAAATGCTCTGTACATGTGTGTACATATT 924

Qy 2276 AGAAAGGGAAGATGTTAAGATAGTGGCTGTGGGTTACACAGGGTGCCTGCCAGCGGTAA 2335
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 925 AGAAAGGGAAGATGTTAAGATAGTGGCTGTGGGTTACACAGGGTGCCTGCCAGCGGTAA 984

Qy 2336 TATATTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTA 2395
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 985 TATATTTAGAAATAATATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTA 1044

Qy 2396 ATTTTTTTCTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTT 2455
        ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 1045 ATTTTTTTCTTTTAAAGAAAGCAGGCTTTTCTAGACTTTAAAGACTAAAGTCTTT 1104

Qy 2456 GGGAGGTCACGTGTAGAGAGAGGAGCTTGAGGCCACCCGACACAAATTCACCCAGAGG 2515
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db 1105 GGGAGGTCACGTGTGTAGAGAGAGGAGCTTGAGGCCACCCGACACAAATTCACCCAGAGG 1164

Qy 2516 GAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGATCACCTGTGTATGTCAACAGAAGG 2575
        |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db 1165 GAAATCTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAG 1224

Qy 2576 GGATACCGTCTTCCTTGAAGAGGAAAACTCTGTCACTCCTCATGCCTCTGTCAGCTCATCACAC 2635
        ||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db 1225 GGATACCGTCTCCTTGAAGAGGAAAACTCTGTCACTCCTCATGCCTCGTCAGCTCATCACAC 1284

Qy 2636 CCATTTCTTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATATCTGCTATATAATTCTC 2695
        ||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db 1285 CCATTTCTTCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTC 1344

Qy 2696 TGTCTCTCTGTTTATCTCTCCCCTTCTTCTCTCCATCTCCATTCTT 2755
        ||||||||| |||||||||||||||||||||||||||||||||||||
Db 1345 TGTCTCTCTG-TTATCTCTCCCCTTCTTCTCTCCATCTCCATTCTT 1403

Qy 2756 TTGAATTTCCTCATCCCTCCAATCCGATCATCCGCAAAAAGGAACAAAAGCAAAAGCACAACACAAACCA 2813
        |||||||||||||||||||||||||||||||||||||||||||||||| ---|||||||||||||||||
Db 1404 TTGAATTTCCTCATCCCTCCAATCCGATCCGATCATCCGACACCCCCCCCAGG 1463

Qy 2814 CAAAGCAGTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCAAAAGCACAACACAAACCA 2873
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1464 CAAAGCAGTGCTCTGAGTATCACATCACACAAAGGAACAAAAGCAAAAGCACAACACAAACCA 1523

Qy 2874 GCCTCAACTTACACTTGGTTACTCAAAGACAAGAGTCAATGCATCAATGTGTCCTAGCGTT 2933
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1524 GCCTCAACTTACACTTGGTTACTCAAAGACAAGAGTCAATGCATCAATGTGTCCTAGCGTT 1583

Qy 2934 TTGGAAGAGGAAAAACAAGGAACCACCAAACCAAACAATCAACCAATCAACAAAGAAAAATTC 2993
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1584 TTGGAAGAGGAAAAACAAGGAACCACCAAACCAAACAATCAACCAATCAACAAAGAAAAATTC 1643

Qy 2994 CACAATGAAAGAATGATATTTTGTCTTTTGCATTTGTCTTTTGGTGTATAAGCCATCAATATTCAG 3053
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1644 CACAATGAAAGAATGATATTTTGTCTTTTGCATTTGTCTTTTGGTGTATAAGCCATCAATATTCAG 1703

Qy 3054 CAAAATGATTCCTTTCTTT-AAAAAAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGT 3112
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1704 CAAAATGATTCCTTTCTTTTAAAAAAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCAAGGT 1763

Qy 3113 TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTAACGAGAAAAACACACAGAAGAAGC 3172
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1764 TGTTGGCCCAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGC 1823

Qy 3173 TACCTCAGGTGTTTTACCTCAGCCTTCAGCCATCCACTTCACCTCGGATCTC 3232
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1824 TACCTCAGGTGTTTTTACCTCAGCCTTCAGCCATCCACTTCACCTCGGATCTC 1883

```
Qy   3233 GCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGA 3292
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1884 GCTGATAGTTGGAGCATTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGA 1943

Qy   3293 TATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGT 3352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1944 TATCAACTGCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGT 2003

Qy   3353 CTTCCGCTAGCCAAGAACNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   2004 CTCCCGCTAGCCAAGAACAACCTATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 2063
```

SQ   Sequence 2010 BP; 565 A; 552 C; 532 G; 361 T; 0 U; 0 Other;

Query Match   58.0%;   Score 1979.8;   DB 5;   Length 2010;
Best Local Similarity   99.4%;   Pred. No. 0;
Matches 1987;   Conservative   0;   Mismatches 12;   Indels 0;   Gaps   0;

```
Qy   35  GCCGGGGAGCCGCGGCCTCCGGGAGAGACGGATGAGACAAGCTTACATCGGGA 94
         ||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db   12  GCCGTAGGGAGCCGCGGCCTCAGGGGAGAGACGGATGAGACAAGCTTACATCGGGA 71

Qy   95  ACCTGAGCCCGCCGTCACCGCCGCCGTCCGCCAGCTCTTTGGGGACAGGAGAGCTGC 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72  ACCTGAGCCCGCCGTCACCGCCGCCGTCCGCCAGCTCTTTGGGGACAGGAGAGCTGC 131

Qy   155 CCCTGGCCGGGACAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCGACCAGA 214
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 CCCTGGCGGGGACAGGTCCTGCTGAAGTCCGGCTACGCCTTCGTGGACTACCCGACCAGA 191

Qy   215 ACTGGGCCCATCCGCGCCATGCCAATGCGAGACCCCTCGGGTAAAGTGGAATTGCATGGGAAAATCA 274
         |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 ACTGGGCCCATCCGCGCCATGCAGACCCCTCGGGTAAAGTGGAATTGCATGGGAAAATCA 251

Qy   275 TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   252 TGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAATTCAGATTCGAAACA 311

Qy   335 TCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTGG 394
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312 TCCCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTGG 371

Qy   395 AGAATGTGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAA 454
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372 AGAATGTGAACAAGTCAACACAGACACAGAAACCGCCGTTGTCAACGTCACATATGCAA 431

Qy   455 CAAGAGAGAAGCAAAAATAGCCATGGGAGGCATCAGTTTGAGAACTACT 514
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   432 CAAGAGAAGAGCAAAAATAGCCATGGGAGGCATCAGTTTGAGAACTACT 491

Qy   515 CCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAGCTCCCCTCGCCCCCTCAGCGAG 574
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   492 CCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAGCTCCCCTTCGCCCCCTCAGCGAG 551

Qy   575 CCCAGCGTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCGGGGCACTTCTCAGG 634
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   552 CCCAGCGTGGGGACCACTCTTCCCGGGAGCAAGGCCACGCCCGGGGCACTTCTCAGG 611

Qy   635 CCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCA 694
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   612 CCAGACAGATTGATTTCCCGCTGCGGATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCA 671

Qy   695 TCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCGGGTAGATA 754
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672 TCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGACCCAGTCCGGGTAGATA 731

Qy   755 TCCATAGAAAAGAGAACTTCTGGAGCTGCGAGAGAAGCTGTCACCATCCATGCCACCCCAG 814
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   732 TCCATAGAAAAGAGAACTTCTGGAGCTGCGAGAGAAGCCTGCTGTCACCATCCATGCCACCCCAG 791

Qy   815 AGGGGACTTCTGAGCATGCCGCATGATCTTGAAATCATGCAGAAGAGGCAGATGAGA 874
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   792 AGGGGACTTCTGAGCATGCCGCATGATCTTGAAATCATGCAGAAGAGGCAGATGAGA 851

Qy   875 CCAAACTAGCCGAAGAGATTCCTCGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC 934
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   852 CCAAACTAGCCGAAGAGATTCCTCGAAAATCTTGGCACACAATGGCTTGGTTGGAAGAC 911

Qy   935 TGATTGGAAAAGAAGGGCAGAAAATTTGAAGAAATGAACATGAAACAGAGGACCAAGATAA 994
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   912 TGATTGGAAAAGAAGGGCAGAAAATTTGAAGAAATGAACATGAAACAGAGGACCAAGATAA 971

Qy   995 CAATCTCATCTTTGCAGGATTTGAGCATATACACCCCGAAAAGAACCATCACTGTGAAGG 1054
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972 CAATCTCATCTTTGCAGGATTTGAGCATATACACCCCGAAAAGAACCATCACTGTGAAGG 1031
```

RESULT 9
AAS70981
ID   AAS70981 standard; cDNA; 2010 BP.
XX
AC   AAS70981;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6785.
XX
KW   Human; chromosome mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06794.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 6785; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX

```
XX
AC   ACA90176;
XX
DT   10-JUL-2003  (first entry)
XX
DE   cDNA encoding novel human protein NOV14a.
XX
KW   Human; Cytostatic; DAPK3-Agonist; DAPK3-Antagonist; cancer; NOV; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200301571-A2.
XX
PD   17-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031357.
XX
PR   05-OCT-2001; 2001US-0327454P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0328029P.
PR   09-OCT-2001; 2001US-0328056P.
PR   12-OCT-2001; 2001US-0328849P.
PR   15-OCT-2001; 2001US-0329414P.
PR   17-OCT-2001; 2001US-0330142P.
PR   22-OCT-2001; 2001US-0341058P.
PR   24-OCT-2001; 2001US-0343629P.
PR   29-OCT-2001; 2001US-0349575P.
PR   01-NOV-2001; 2001US-0346357P.
PR   25-JUN-2002; 2002US-0391342P.
PR   01-OCT-2002; 2002US-0262445.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Alsobrook JP, Burgess CE, Catterton E, Chant JS, Chaudhuri A;
PI   Edinger SR, Gerlach VL, Giot L, Gorman L, Guo X, Kekuda R;
PI   Mezes PS, Millet I, Ooi CE, Patturajan M, Rieger DK, Spytek KA;
PI   Taupier RJ, Zerhusen BD, Zhong H, Zhong M;
XX
DR   WPI; 2003-381704/36.
DR   P-PSDB; ABU89799.
XX
PT   New DAPK3 polypeptide, useful for preparing a composition for treating or
PT   preventing e.g., cancer.
XX
PS   Claim 20; Page 129; 253pp; English.
XX
CC   The invention describes an isolated polypeptide comprising any of 33 90-
CC   1273 amino acid sequences (I) given in the specification or its mature
CC   form, a sequence that is at least 95 % identical to (I), or a sequence
CC   comprising one or more conservative substitutions in the amino acid
CC   sequence of (I). The polypeptide is useful for preparing a composition
CC   for treating or preventing e.g. cancer. This sequence encodes a novel
CC   human NOV protein
XX
SQ   Sequence 1707 BP; 492 A; 443 C; 442 G; 330 T; 0 U; 0 Other;

Query Match       41.7%;  Score 1423.8;  DB 7;  Length 1707;
Best Local Similarity   88.4%;  Pred. No. 0;
Matches 1625;  Conservative   0;  Mismatches   82;  Indels  132;  Gaps   2;
```

```
Qy   1055  GCACAGTTGAGGCCTGCTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCT  1114
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1032  GCACAGTTGAGGCCTGCTGCCAGTGCTGAGATAGAGATTATGAAGAAGCTGCGTGAGGCCT  1091

Qy   1115  TTGAAAATGATATGCTGGCTGTTAACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCA   1174
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1092  TTGAAAATGATATGCTGGCTGTGTTAACCAACAAGCTGATCCCAGGGTTGAACCTCA     1151

Qy   1175  GCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCG   1234
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152  GCGCACTTGGCATCTTTCAACAGGACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCG   1211

Qy   1235  GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCC  1294
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212  GAGCTCCCCCGCTGCCCCCTACCACCCCTTCACTACCCCACTCCGGATACTTCTCCAGCC  1271

Qy   1295  TGTACCCCCATCACCAGTTGGCCCGTTCCCGCCGCATCATCATCTTATCCAGAGCAGGAGA  1354
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272  TGTACCCCCATCACCAGTTGGCCCGTTCCCGCCGCATCATCATCTTATCCAGAGCAGGAGA  1331

Qy   1355  TTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCCATCATGGGAAGAAGGGGCAC    1414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332  TTGTGAATCTCTTCATCCCAACCCAGGCTGTGGGCGCCCATCATGGGAAGAAGGGGCAC    1391

Qy   1415  ACATCAAACAGCTGGCGAGATTCGCCGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCC  1474
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392  ACATCAAACAGCTGGCGAGATTCGCCGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCC  1451

Qy   1475  CAGAGCTCAGCGGAAAGGATGGTCATCATCACCGGCCACCGGAGCCCCAGTTCAAGGCCC  1534
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452  CAGAGCTCAGCGGAAAGGATGGTCATCATCACCGGCCACCGGAGCCCCAGTTCAAGGCCC  1511

Qy   1535  AGGGACGGACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAGAAGTGA  1594
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512  AGGGACGGACGGATCTTTGGGAAACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAGAAGTGA  1571

Qy   1595  AGCTGGAAGCCATATCAGAGTGCCCTCTTCCACAGTCGCCAGTGCCGGTGATTGGCAAAGTG  1654
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572  AGCTGGAAGCCATATCAGAGTGCCCTCTTCCACAGTCGCCAGTGCCGGTGATTGGCAAAGTG  1631

Qy   1655  GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACC  1714
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632  GCAAGACCGTGAACGAACTGCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACC  1691

Qy   1715  AAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTCTTTGCTAGCC   1774
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692  AAACGCCAGATGAAAATGAGGAAGTGATCGTCAGAATTATCGGGCACTCTTTGCTAGCC   1751

Qy   1775  AGACTGCCACAGCGCAAGATCAGCGGAAATTGTACAACAGTGAAGCAGCAGGAGCAGAAAT  1834
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752  AGACTGCCACAGCGCAAGATCAGCGGAAATTGTACAACAGTGAAGCAGCAGGAGCAGAAAT  1811

Qy   1835  ACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCACAGCAAAACA  1894
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812  ACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGCTCCCACAGGCACCACAGCAAAACA  1871

Qy   1895  ACGGATGAATGATAGCCCTGACACACCTGACAGAATGAGACAAACGCAGCCAGCCAGAT   1954
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872  ACGGATGAATGTGCCCTTCCAACACCTGACACCTGACAGAATGAGACAAACGCAGCCAGCAT  1931

Qy   1955  CGGGAGCAAACCAAAGACCAATCTGAGGAATGAGAGTCTGCGGAGGCCGGCCAGGGACTCT  2014
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932  CGGGAGCAAACCAAAGACCAATCTGAGGAATGAGAGTCTGCGGAGGCCGGCCAGGGACTCT  1991

Qy   2015  GCCGAGGGCCCTGAGAACCC  2033
           ||||||||||||||||||||
Db   1992  GCCGAGGGCCCTGAGAACCC  2010
```

```
RESULT 10
ACA90176
ID   ACA90176 standard; cDNA; 1707 BP.
```

```
Qy    246  GGGTAAGTGGAATTGCATGGGAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCT  305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GGGTCAAGTGCAATTGCATGGGAAATCATGGAAGTTGATTATTCAGTCTCTATAAAGCT  240

Qy    306  AAGGAGCAGGAGAAATTCAGATTGAAACATCCCTCCTCACCTGCAGTTGGGAGTGTTGGA  365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  AAGGAGCAGGAGACAATTCCGATTCGAAATATCCCTCCCTCACCTGCAGTGGGAGTGTTGGA  300

Qy    366  TGGACTTTTGGCTCAATATGGGAACAGTGGACAGTGGAACAAGTCAACACAGACACAGA  425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  TGGACTTTTGGCTCAATATGGGAACAGTGGACAGTGGAACAAGTCAACACAGACACAGA  360

Qy    426  AACCCGTTGTCAAGTCACATATGCAACAAGAGAAGAAAATAGCCATGGAGAA  485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GACCCGTTGTTGTCAAGTCACATATGCAACAAGAGAAGAAGTAAAAATAGCCATGAAGAA  420

Qy    486  GCTAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGA  545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  GCTAAGCGGGGCATCAGTTTGAGAACCACTACTTCAAGATTTCCTACATCCCGGATGACGA  480

Qy    546  GGTGAGCTCCCCTTCGCCCCTCAGCGTGGGGCACTTCCAGCCCACCTCTCCCGGGACCA  605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GGTGAGCTCGCCCTTCGCGCCCTCAGCGCGAGCCCAGCCCACCTCTTTCCTGGGACCA  540

Qy    606  AGGCCACGCGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCCGCTGCGGATCCT  665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AGGCCAAGCGCCCTGGGGGCTCTTCTCAGGCCAGACAGATTGATTTCCCACTGCGTGTCCT  600

Qy    666  GGTCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAACAT  725
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GTTCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAGGAGGGCTTGACCATAAAGAACAT  660

Qy    726  CACTAAGCAGACCCAGTCCCGGTAGTAATCTATAGACAAGAGACTCGGAGTGCGACGA  785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CACTAAGCAGACGAGCCCGTCCCGGGTAGACATCTATAGACAAGAGACTCTCCAGAGTGCGACGA  720

Qy    786  GAAGCCTGTCACCATCCATGCCCAGAGGGGACTTCTGAAGCATGCCGATGATTCT  845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  GAAGCCTGTCACCATGCCATGCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCT  780

Qy    846  TGAAATCATGCAGAAAGAGAGGCAGTTGCAGATGAGACCCAAACTAGCCGAGAGATTCCTCTGAAAAT  905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  TGAAATAATGCAGAAAGAGAGGCAGTTGCAGATGAGACCCAAACTAGCCGAGAGATTCCTCTGAAAAT  840

Qy    906  CTTGGACAACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGAGGCAGAAATTTGAAGAA  965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CTTGGCCCACAATGCTTGGTTGGAAGACTGATTGGAAGAAGAGGCAGAAATTTGAAGAA  900

Qy    966  AATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCGATGATTTGAGCATATA  1025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  AAATGAACATCGAAACAGGGACCAAGATAACAATCTCATCTTTCGCGATGATTTGAGCATATA  960

Qy    1026 CAACCCGGAAAGAACAACCATCACTGTGAAGGGCACAGTTGAGCGCCTGTGCCAGTGCTGAGAT  1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  CAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGTGCGAGGTGCTGTGCCAGTGCTGAGAT  1020

Qy    1086 AGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAATGATATGTGGCTGTGTTAACCAACA  1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAATGATACGCTGACTGTTAA------  1074

Qy    1146 AGCCAATCTGATCCCCAGGGTTGAACCTTGGCATCTTTTCAACAGGACTGTC  1205
           ||||||||||||||||||||||||||||
Db    1075 -------------------------------------------  1074

Qy    1206 CGTGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCGCCCCTACCACCCCT  1265
           ||||||||||||||||||||||||||||||||||||||||||||
Db    1075 -------------------------------------------  1074

Qy    1266 CACTACCCACTCCGGATACTTCTCGTACCCCATCACCAGTTTGGCCCGTTCCC  1325
           ||||||||||||||||||||||||||||||||||||||||
Db    1075 ---TACCCACTTCGGATACTTCTCCCAGCCTGTACCCCCATGCCCAGTTTGGCCCGTTCCC  1131

Qy    1326 GCATCATCACTCTTATCTCCAGAGCAGAGCAGAGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGT  1385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1132 GCATCATCACTCTTATCCAGAGCAGCAGAGGATTGTCAATCTCTTCATCCCAACCCAGGGTGT  1191

Db    1386 GGGCGCCATCATCGGGAAGAAGGGGCACAGCAGGAGATTGTCAATCTCTTCATCCCAACAGTCGGCGAGATTGCCCGGAGC  1445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1192 GGGCGCCATCATCGGGAAGAAAGGGCACACATCAACAGCTGCGAGATTCGTGGGGAGC  1251

Qy    1446 CCTATCAAGATGCCCCTGCGGAAGGCCCAGACGTCAGCGAACGAATGGTCATCATCAC  1505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1252 CTCCATCAAGATCGCCCCTGCGGAGATCCCC---CCTGCGCTCAGCGGAAGGTCATCATCAC  1308

Qy    1506 CGGGCCACCGGAAGCCCAGTTCAGGCCCCAGGGACCAGGACGGATCTTTGGGAAACTGAAAGAGGA  1565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1309 CTGCCCACCGGAAGCCCAGTTCAGGCCCCAGGGACCAGGACGGATCTTTGGGAAACTGAAAGAAGA  1368

Qy    1566 AAACTTCTTAACCCCAAAGAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCCTTC  1625
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1369 AAACTTTTTTAACCCCAAAGAACGTGAAGCTGGAAACCCATATCAGAGTGCCCTCTTC  1428

Qy    1626 CACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGACGAACTGCAGAACTAAC  1685
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1429 CACCGCTGCGGGGTGATTGGCAAAGGGGGCAAGACCGTGAATGAACTGCAGAATTTAAT  1488

Qy    1686 CAGTGCAGAAGTCATCGTCGTCCTCGTGCCAGTGGAAATGAAAATGAGGGAAGTGATCGT  1745
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1489 CAGTGCAGAAGTCATCGTCGTCCTCGTGCCAGTGGAAATGAAAATGAGGGAAATGATCGT  1548

Qy    1746 CAGAATTATCGGGCACTCTTTTGCTAGCCAGACGAGAAATACCCTCAGGGAATCAGGGAAATTGT  1805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1549 CAGAATTATCGGGCACTTCTTTGCTAGCCAGACTCAGCGAGAAATACCCTCAGGGAATCAGGGAAATTGT  1608

Qy    1806 ACACAGTGAAGCAGCAGGCACCAGCAAAACAACGGATGAAT  1865
           ||||||||||||||||||||||||||||||||||||||||
Db    1609 ACACAGTGAAGCAGCAGGCACCAGCAAAACAACGGCAGCAGCAA  1668

Qy    1866 GTGAGGCTCCCACAGGCCACCAGCAAAACAACGGATGAAT  1904
           ||||||||||||||||||||||||||||||||||||||
Db    1669 GTGAGGATCCCACAGGCCACCAGGCCACAAGCAAACAACGGAAGAAT  1707
```

RESULT 11
AAS26570
ID AAS26570 standard; cDNA; 1186 BP.
XX
AC AAS26570;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 749.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cereboprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.

```
PR   17-MAR-2000;   2000US-0190076P.
PR   18-APR-2000;   2000US-0198123P.
PR   19-MAY-2000;   2000US-0205515P.
PR   07-JUN-2000;   2000US-0209467P.
PR   28-JUN-2000;   2000US-0214886P.
PR   30-JUN-2000;   2000US-0215135P.
PR   07-JUL-2000;   2000US-0216647P.
PR   07-JUL-2000;   2000US-0216880P.
PR   11-JUL-2000;   2000US-0217487P.
PR   11-JUL-2000;   2000US-0217496P.
PR   14-JUL-2000;   2000US-0218290P.
PR   26-JUL-2000;   2000US-0209963P.
PR   26-JUL-2000;   2000US-0209964P.
PR   14-AUG-2000;   2000US-0224518P.
PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   25-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   26-SEP-2000;   2000US-0234998P.
PR   27-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   29-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.

PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0244474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   01-DEC-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   05-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   06-DEC-2000;   2000US-0256719P.
PR   08-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   11-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0256678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16583.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 749; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
```

CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the proteins and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation.
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

```
Query Match             32.1%;  Score 1096.8;  DB 4;  Length 1186;
Best Local Similarity   98.4%;  Pred. No. 1.3e-280;
Matches  1145;  Conservative   3;  Mismatches  12;  Indels   4;  Gaps   4;

Qy  1376 CCCAGGCTGTGGGCGCCCATCATCGGCGAAGAAGGGGCCACACATCAAAACAGTCGCGAGAT 1435
         ||||  ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    25 CCCACGCGTCCGGCGCCATCATCGGCGAAGAAGGGGCCACACATCAAAACAGTCGCGAGAT   84

Qy  1436 TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAAGGATGG 1495
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    85 TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAAGGATGG  144

Qy  1496 TCATCATCACCGGCGCCGGAAGCCACCGGAAGCCAGTTCAAGGCCCAGTTCAAGGCGGAAAC 1555
         |||||||||||||||||  ||: |||||||||||||||||||||||||||||||||||||
Db   145 TCATCATCACCGGCGCCA-CGSAAGCCCAGTTCAAGGCCCAGTTCAAGGCGGATCTTTGGGAAAC  203

Qy  1556 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAGAGTGAAGCTGAAGGCCATATCAGAG 1615
         |||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db   204 TGAAAGAGGAAAACTTCTTTAACCCCAAAGAGAGTGAAGCTGAACCCATTATCAGAG  263

Qy  1616 TGCCCTCTTCCACAGCTGGCCGGGTGATTGCCAAAGGTGGCAAAGACCGTGAACGAACTGC 1675
         ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264 TGCCCTTCTTCCACAGCTGGCCGGGTGATTGCCAAAGGTGGCAAAGACCGTGAACGAACTGC  323

Qy  1676 AGAACTTAACCAGTGCAGAAGTCGCAGAAGTGAGAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG 1735
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   324 AGAACTTAACCAGTGCAGAAGTCGCAGAAGTGAGAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGG  383

Qy  1736 AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA 1795
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   384 AAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCA  443

Qy  1796 GGGAAATTG-TACACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTCA 1854
         |||||||||  ||:|||||||||||||||||||||||||||||||||||||||||||
Db   444 GGGAAATTGTTACAACAGGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCCTCA  503

Qy  1855 CAGCGCGCAAGCAAGTGAGGCTCCACAGGCACCAGAACACAACGGATGAATGTAGCCCTTC 1914
         |||||||||  |||||||||||||| ||| ||||  ||| |||||||||||||||||||||
Db   504 CAGCGCGCA-NAAGTGAGGCTCCCACACGCACCAAACCAGCACAAACAACGGATGAATGTAGCCCTTC  562

Qy  1915 CAACACCTGACAGAGAATGAGACCAAACGCAGCCAGATCGGAGGAGCAAACCAAAGACCA 1974
         |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   563 CAACACCTGACAGAGAATGAGACCAAACGCAGCCAGATCGGGAGGAGCAAACCAAAGACCA  622

Qy  1975 TCTGAGGAATGAGAAGTCTGCGGAGGCGGCCCAGGAGGCTCGCCGAGGCCCTGAGAACCCC 2034
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623 TCTGAGGAATGAGAAGTCTGCGGAGGCGGCCCAGGAGGCTCGCCGAGGCCCTGAGAACCCC  682

Qy  2035 AGGGGCCGAGGAGGGGGCGGGGAAGGTCAGCCAGTTGCCAGAACCACCGGAGCCCGCGCCT 2094
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   683 AGGGGGCCGAGGAGGGGGCGGGGAAGGTCAGCGGGGAAGGTTGCCAGAACCACCGGAGCCCGCGCCT  742

Qy  2095 CCCGCCCCCAGGGCTTCTGCGAGGCTTCAGCCAT-CCACTTCACCATCCACTCGGATCTC 2153
         |||||||||  ||||||||||||||||||||||| |||||||||||||||||||||||||
Db   743 CCCGCCCCCCAGGGCTTCTGCGAGGCTTCAGCNATNCCACTTCACCATCCACTCGGATCTC  802

Qy  2154 TCCTGAACTCCCACGAGGCTATCCCTTTAGTTGAACTAACATAGGTGAACCGTGTTCAAA 2213
         |||||||||||||||| |||||||||||||||||||||||||||||||||  |||||||||
Db   803 TCCTGAACTCCCACGACGGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAA  862

Qy  2214 GCCAAGCAAAATGCACACCTTTTCTGTGGCAAATCGTCTGTACACATGTGTGTACATA 2273
         |||||||||||||||||||||||||||||||||||||  ||||||||||  |||||||
Db   863 GCCAAGCAAAATGCACACCCTTTTCTGTGGCAAATGTCTCTGTACACAGGGTGCTGTACATA  922

Qy  2274 TTAGAAAGGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGGTGCCTGCAGCGGT 2333
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   923 TTAGAAAGGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGT  982

Qy  2334 AATATATTTTAGAAATAATATATCAAATAACTCCAATTTTAATCAATTAT 2393
         |||||||||||||||||||||||||||||||||  ||||||||||||||||
Db   983 AATATATTTTAGAAATAATATATCAAATAACTAACTCCAATTTTAATCAATTAT 1042

Qy  2394 TAATTTTTTTTTCTTTTAAAGAGAAAGCAGGCTTTTCTAGACTTTAAAGATAAAGTCT 2453
         ||||||||||  |||||: ||||||||||: |||||||||||||||||||| |||||||
Db  1043 TAATTTTTTTTTTCTTTTTAAAGAGRAAGCMGGCTTTTCTAGACTTTAAAGAATAAAGTCT 1102

Qy  2454 TTGGGAGGTTCTCACGGTGTAGAGAGACGTTTGAGGCCACCCGCACAAAATTCACCCAGA 2513
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1103 TTGGGAGGTTCTCACGGTGTAGAGAGACGTTTGAGGCCACCCGCACAAAATTCACCCAGA 1162

Qy  2514 GGGAAATCTCGTCGGAAGGACACT 2537
         ||||||||||||||||||||||||
Db  1163 GGGAAATCTCGTCGGAAGGACACT 1186
```

RESULT 12
ABX73911
ID   ABX73911 standard; DNA; 1186 BP.
XX
AC   ABX73911;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human novel polynucleotide #739.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.

```
Query Match            32.1%;  Score 1096.8;  DB 7;  Length 1186;
Best Local Similarity  98.4%;  Pred. No.1.3e-280;
Matches 1145;  Conservative 3;  Mismatches 3;  Indels 4;  Gaps 4;

Qy   1376 CCCAGGCTGTGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGAT 1435
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 CCCACGGCTCCGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGAT 84

Qy   1436 TCGCCGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCGAAGCGTCAGCGAAAGGATGG 1495
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 TCGCGGGAGCCTCTATCAAGATTGCCCCTGCCGAAGGCCCGAAGCGTCAGCGAAAGGATGG 144

Qy   1496 TCATCATCACCGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTTGGGAAAAC  1555
               ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    145 TCATCATCACCGGCCA-CGGAAGCCCAGTTCAAGGCCCAGGGACCGGATCTTTGGGAAAAC  203

Qy   1556 TGAAAGAGGAAAACTTCTTAACCCCAAAGAGAAGAAGTGAAGCTGGAAGCGCATATCAGAG 1615
               |||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db    204 TGAAAGAGGAAAACTTCTTAACCCCAAAGAGAAGAAGTGAAGCTGGAACCCATTATCAGAG 263

Qy   1616 TGCCCTCTTCCACAGCTGGCCGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGC     1675
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    264 TGCCCTCTTCCACAGCTGGCCGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGC     323

Qy   1676 AGAACTTAACCAGTGCAGAATTATCGGGCACTCTTCTTGCTAGCCAGACTGCACAGCGCAAGATCA 1735
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 AGAACTTAACCAGTGCAGAATTATCGGGCACTCTTCTTGCTAGCCAGACTGCACAGCGCAAGATCA 383

Qy   1736 AAGTGATCGTCAGAATTATCGGGCACTCTTCTTGCTAGCCAGACTGCACAGCGCAAGATCA 1795
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    384 AAGTGATCGTCAGAATTATCGGGCACTCTTCTTGCTAGCCAGACTGCACAGCGCAAGATCA 443

Qy   1796 GGGAAATTG-TACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCA 1854
               |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    444 GGGAAATTGTTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTGCGCCTCA 503

Qy   1855 CAGGCGCAGTGAAGTGAGGCTCCCACAGGCACCAGCAGGCACCAGATGAATGTAGCCCTTC 1914
               ||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    504 CAGGCGCA-NAAGTGAGGCTCCCACAGGCACCAGCAGGCACCAGATGAATGTAGCCCTTC  562

Qy   1915 CAACACCTGACAGAATGAGACCAAACAGCAGCCAGCCCAGATCGGGAGCAAACCAAAGACCA 1974
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 CAACACCTGACAGAATGAGACCAAACAGCAGCCAGCCCAGATCGGGAGCAAACCAAAGACCA 622

Qy   1975 TCTGAGGAATGAGAAGTCTCGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCG 2034
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    623 TCTGAGGAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGCCCTGAGAACCCG  682

Qy   2035 AGGGGCGGAGGAGGGGGCGGGGAAGGTCAGCAGGTTTGCCAGAACGACCACCCAGCCCGCT 2094
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    683 AGGGGGCGGAGGAGGGGGCGGGGAAGGTCAGCAGGTTTGCCAGAACGACCACCCAGCCCGCT 742

Qy   2095 CCCGCCCCCCAGGGCTTCTGCAGGCTTCAGCCAT-CCACTTCACCATCCACTTCACCATCCGGATCTC 2153
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    743 CCCGCCCCCCAGGGCTTCTGCAGGCTTCAGCNATNCCACTTCACCATCCACTTCACCATCCGGATCTC  802

Qy   2154 TCCTGAACTCCCACGCGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAA      2213
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    803 TCCTGAACTCCCACGACGCTATCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAA      862

Qy   2214 GCCAAGCAAAATGCACACCCTTTTCGTGGCAAATGCTCTCGTACATGTGTGTACATA 2273
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 GCCAAGCAAAATGCACACCCTTTTTCGTGGCAAATGCTCTCGTACATGTGTGTACATA 922

Qy   2274 TTAGAAGGAAGATGTTAAGATATGTGGCCTGTGGGTTACAACAGGGTGCCTGAGCCGGT 2333
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    923 TTAGAAGGAAGATGTTAAGATATGTGGCCTGTGGGTTACAACAGGGTGCCTGAGCCGGT 982

Qy   2334 AATATATTTTAGAAATAAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTAT 2393
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    983 AATATATTTTAGAAATAATATATCAAATAACTAACTCAACTAACTCCAATTTTTAATCAATTAT 1042
```

```
Qy    2394 TAATTTTTTTCTTTTAAAGAGAATAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCT 2453
           ||||||||| |||||  ||||||| |||| |||||||||||||||||||||||||||
Db    1043 TAATTTTTTTCTTTTTAAAGAGRAAGCMGGCTTTCTAGACTTTAAAGAATAAAGTCT 1102

Qy    2454 TTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCGCACAAAATTCACCCAGA 2513
           |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    1103 TTGGGAGGTCTCACGGTGTAGAGAGGAGCTTTGAGGCCACCCCGCACAAAATTCACCCAGA 1162

Qy    2514 GGGAAATCTCGTCGGAAGGACACT 2537
           ||||||||||||||||||||||||
Db    1163 GGGAAATCTCGTCGGAAGGACACT 1186
```

RESULT 13
AAS26151
ID   AAS26151 standard; cDNA; 822 BP.
XX
AC   AAS26151;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 330.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO20015322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX

PR 31-JAN-2000; 2000US-0179065P.	PR 22-AUG-2000; 2000US-0226868P.
PR 04-FEB-2000; 2000US-0180628P.	PR 22-AUG-2000; 2000US-0227182P.
PR 24-FEB-2000; 2000US-0184664P.	PR 23-AUG-2000; 2000US-0227009P.
PR 02-MAR-2000; 2000US-0186350P.	PR 30-AUG-2000; 2000US-0228924P.
PR 16-MAR-2000; 2000US-0189874P.	PR 01-SEP-2000; 2000US-0229287P.
PR 17-MAR-2000; 2000US-0190076P.	PR 01-SEP-2000; 2000US-0229343P.
PR 18-APR-2000; 2000US-0198123P.	PR 01-SEP-2000; 2000US-0229344P.
PR 19-MAY-2000; 2000US-0205515P.	PR 01-SEP-2000; 2000US-0229345P.
PR 07-JUN-2000; 2000US-0209467P.	PR 05-SEP-2000; 2000US-0229509P.
PR 28-JUN-2000; 2000US-0214886P.	PR 05-SEP-2000; 2000US-0229513P.
PR 30-JUN-2000; 2000US-0215135P.	PR 06-SEP-2000; 2000US-0230437P.
PR 07-JUL-2000; 2000US-0216447P.	PR 06-SEP-2000; 2000US-0230438P.
PR 07-JUL-2000; 2000US-0216880P.	PR 08-SEP-2000; 2000US-0231242P.
PR 11-JUL-2000; 2000US-0217487P.	PR 08-SEP-2000; 2000US-0231243P.
PR 11-JUL-2000; 2000US-0217496P.	PR 08-SEP-2000; 2000US-0231244P.
PR 14-JUL-2000; 2000US-0218290P.	PR 08-SEP-2000; 2000US-0231413P.
PR 26-JUL-2000; 2000US-0220963P.	PR 08-SEP-2000; 2000US-0231414P.
PR 26-JUL-2000; 2000US-0220964P.	PR 08-SEP-2000; 2000US-0232080P.
PR 14-AUG-2000; 2000US-0224518P.	PR 08-SEP-2000; 2000US-0232081P.
PR 14-AUG-2000; 2000US-0224519P.	PR 12-SEP-2000; 2000US-0231968P.
PR 14-AUG-2000; 2000US-0225213P.	PR 14-SEP-2000; 2000US-0232397P.
PR 14-AUG-2000; 2000US-0225214P.	PR 14-SEP-2000; 2000US-0232398P.
PR 14-AUG-2000; 2000US-0225266P.	PR 14-SEP-2000; 2000US-0232399P.
PR 14-AUG-2000; 2000US-0225267P.	PR 14-SEP-2000; 2000US-0232400P.
PR 14-AUG-2000; 2000US-0225268P.	PR 14-SEP-2000; 2000US-0232401P.
PR 14-AUG-2000; 2000US-0225270P.	PR 14-SEP-2000; 2000US-0233063P.
PR 14-AUG-2000; 2000US-0225447P.	PR 14-SEP-2000; 2000US-0233064P.
PR 14-AUG-2000; 2000US-0225757P.	PR 14-SEP-2000; 2000US-0233065P.
PR 14-AUG-2000; 2000US-0225758P.	PR 21-SEP-2000; 2000US-0234223P.
PR 14-AUG-2000; 2000US-0225759P.	PR 21-SEP-2000; 2000US-0234274P.
PR 18-AUG-2000; 2000US-0226279P.	PR 25-SEP-2000; 2000US-0234997P.
PR 22-AUG-2000; 2000US-0226681P.	PR 25-SEP-2000; 2000US-0234998P.
	PR 26-SEP-2000; 2000US-0235484P.
	PR 27-SEP-2000; 2000US-0235834P.
	PR 29-SEP-2000; 2000US-0235836P.
	PR 29-SEP-2000; 2000US-0236327P.
	PR 29-SEP-2000; 2000US-0236367P.
	PR 29-SEP-2000; 2000US-0236368P.
	PR 29-SEP-2000; 2000US-0236369P.
	PR 29-SEP-2000; 2000US-0236370P.
	PR 02-OCT-2000; 2000US-0236802P.
	PR 02-OCT-2000; 2000US-0237037P.
	PR 02-OCT-2000; 2000US-0237038P.
	PR 02-OCT-2000; 2000US-0237039P.
	PR 02-OCT-2000; 2000US-0237040P.
	PR 13-OCT-2000; 2000US-0239935P.
	PR 13-OCT-2000; 2000US-0239937P.
	PR 20-OCT-2000; 2000US-0240960P.
	PR 20-OCT-2000; 2000US-0241221P.
	PR 20-OCT-2000; 2000US-0241785P.
	PR 20-OCT-2000; 2000US-0241786P.
	PR 20-OCT-2000; 2000US-0241787P.
	PR 20-OCT-2000; 2000US-0241808P.
	PR 20-OCT-2000; 2000US-0241809P.
	PR 20-OCT-2000; 2000US-0241826P.
	PR 01-NOV-2000; 2000US-0244617P.
	PR 08-NOV-2000; 2000US-0246474P.
	PR 08-NOV-2000; 2000US-0246475P.
	PR 08-NOV-2000; 2000US-0246476P.
	PR 08-NOV-2000; 2000US-0246477P.
	PR 08-NOV-2000; 2000US-0246478P.
	PR 08-NOV-2000; 2000US-0246523P.
	PR 08-NOV-2000; 2000US-0246524P.
	PR 08-NOV-2000; 2000US-0246525P.
	PR 08-NOV-2000; 2000US-0246526P.
	PR 08-NOV-2000; 2000US-0246527P.
	PR 08-NOV-2000; 2000US-0246528P.
	PR 08-NOV-2000; 2000US-0246532P.
	PR 08-NOV-2000; 2000US-0246609P.
	PR 08-NOV-2000; 2000US-0246610P.
	PR 08-NOV-2000; 2000US-0246611P.
	PR 08-NOV-2000; 2000US-0246613P.
	PR 17-NOV-2000; 2000US-0249207P.

```
PR   17-NOV-2000;  2000US-0242208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0256678P.
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16164.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 330; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
```

```
Qy    779  CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA   838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA    60

Qy    839  TGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTC    898
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TGATTCTTGAAATCATGCAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTC    120

Qy    899  TGAAAATCTTGGCACACAATGGCTTGGTTGGTGGAAGACTGATTGGAAAGAAGGCAGAAATT   958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TGAAAATCTTGGCACACAATGGCTTGGTTGGTGGAAGACTGATTGGAAAAGAAGGCAGAAATT  180

Qy    959  TGAAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTGAGGGATTTGA    1018
           ||||||||||||||||||||||||||||||||||||||| ||||||:|||||||||||
Db    181  TGAAGAAAATTGAACATGAAACAGGGACCAAGRTAACAATCCATCTTTGAGGGATTTGA    240

Qy   1019  GCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTG   1078
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GCATATACAACCCGGAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTG    300

Qy   1079  CTGAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTA   1138
           ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CTGAGATAGAGATTATGAAGAGCTGCGTGAGGCCTTGAAAATGATATGCTGGCTGTTA    360

Qy   1139  ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAG   1198
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ACCAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAG    420

Qy   1199  GACTGTCCGTGCTATCTCCACCAGCAGCAGGCCCCGGGAGCTCCCCCGCTGCCCCCTACC   1258
           ||||||||||||||||||||||||:||||||||||| :|||||||||||||||||||||:
Db    421  GACTGTCCGTGCTGTATCTCCANCAGCAGCGCCGGGAG-TCCCCCGCTGCCCCCTACS    479

Qy   1259  ACCCCTTCACTACCTCCGGATACTTCTTCCAGCTGCGTACCCCCATCACCAGTTTGGCC   1318
           |:||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db    480  ASCCCTTCACTAMCAMTCCGGATAMTTCTTCCAGCTGCGTACCCCCATCACCAGTTTGGCC    539

Qy   1319  CGTTCCCGCCATCATCACTCTTATCCAGACGAGGAGATTGGAATCTTCATCCCAACCC   1378
           |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db    540  CGTTCCCGCCATCATCACTCTTATCCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCC  599

Qy   1379  AGGCTGTGGGGGCCATCATCGGGAAGAAAGGGGCCACATCAAAACAGCTGGCGAGATTCG   1438
           ||||||||||||||||||||||||||||||||||||||| :|||||||||||||||||||
Db    600  AGGCTGTGGGGGCCATCATCGGGAAGAAGGGGCCACACATCAAAACAGCTGGCGAGATTCG    659

Qy   1439  CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGGTCA   1498
           ||||||||||||||||||||:||||| |||||||||||||||||||||||||||||||||||
Db    660  CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGGTCA    719

Qy   1499  TCATCACCGGGCCACCGGAA-GCCCAGTTCAAGGCCCAGGACGGCGAGATTCTTTGGAAACTG   1557
           ||||||||||||||||||||:||||||||||||||||||||:|||||
Db    720  TCATCAMCGGGCCAMCGGAATGCCCAGTTCAAGGCCCAGGACGGCGATGTTTTTGGGAAYTG    779

Qy   1558  AAAG-AGGAAAACTTCTTAACCCAAAGAAGAGTGAAGCTG   1599
           ||||  ||||||  ||||||:|||||||||||||||||||
Db    780  AAAGAGGAAAATTTTTTAAACCCCAAAGAAGAGTGAAGCTG    822
```

```
RESULT 14
ABX73492
ID   ABX73492 standard; DNA; 822 BP.
XX
AC   ABX73492;
XX
DT   18-MAR-2003   (first entry)
XX
DE   Human novel polynucleotide #320.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
```

```
Query Match      22.7%;  Score 776.2;  DB 4;  Length 822;
Best Local Similarity   97.8%;  Pred. No. 1.9e-195;
Matches  805;  Conservative  10;  Mismatches  10;  Indels  5;  Gaps  3;
```

```
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
PD   19-SEP-2002.
XX
PF   17-JAN-2001;  2001US-00764864.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   28-JUN-2000;  2000US-0214886P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0254447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   22-AUG-2000;  2000US-0226868P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   08-SEP-2000;  2000US-0231413P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0234997P.
PR   27-SEP-2000;  2000US-0235834P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241809P.
PR   01-NOV-2000;  2000US-0244617P.
PR   17-NOV-2000;  2000US-0249299P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
XX
PA   (ROSE/)  ROSEN C A.
PA   (RUBE/)  RUBEN S M.
PA   (BARA/)  BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55232.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
```

```
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 330; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC   human novel polynucleotides of the invention
XX
SQ   Sequence 822 BP; 234 A; 212 C; 201 G; 163 T; 0 U; 12 Other;

Query Match             22.7%;  Score 776.2;  DB 7;  Length 822;
Best Local Similarity   97.8%;  Pred. No. 1.9e-195;
Matches  805;  Conservative  10;  Mismatches  5;  Indels  3;  Gaps  3;

Qy    779 CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA 838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCA 60

Qy    839 TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTC 120

Qy    899 TGAAAATCTTGGCACACAATGGCTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATT 958
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGAAAATCTTGGCACACAATGGCTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATT 180

Qy    959 TGAAGAAAATTGAACATGAACAGGAGACCAAGATAACAATCCATCTTTGCAGGATTTGA 1018
          |||||||||||||||||||||||||||||||||||||: :|||||||||||||||||||
Db    181 TGAAGAAAATTGAACATGAACAGGAGACCAAGRTAACAATCTCATCTTTGCAGGATTTGA 240

Qy   1019 GCATATACAACCCGGAAAGAACAACCATCACTGTGAAGGCACAGTGAGCCTGAGCCTGTGCCAGTG 1078
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCATATACAACCCGGAAAGAACAACCATCACTGTGAAGGCACAGTGAGCCTGAGCCTGTGCCAGTG 300

Qy   1079 CTGAGATAGAGATTATGAAGAAGTCGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTA 1138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGAGATAGAGATTATGAAGAAGTCGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTA 360

Qy   1139 ACCAACAAGCCAATCTGATCCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAG 1198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACCAACAAGCCAATCTGATCCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAG 420

Qy   1199 GACTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCCCCGCTGCCCCCTACC 1258
          |||||||||||||||||||||||||||||||: |||||||||||| : :|||||||
Db    421 GACTGTCCGTGCTATCTCCANCAGCAGGGCCCCGNGGAG-TCCCCCGCTGCCCCCTACS 479

Qy   1259 ACCCCTTCACTACCCACTCCGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCC 1318
          :||: :||: :||||||: : :|||||||||||||||||||||||||||||||||||
Db    480 ASCCCTTCACTAMCCAMTCCGATAMTTCTCCAGCCTGTACCCCATCACCAGTTTGGCC 539

Qy   1319 CGTTCCCGCATCATCACTCTTATCCGAGCAGCAGGAGATTGTGAATCTCTTCTTCATCCCAACCC 1378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 CGTTCCCGCATCATCACTCTTATCCGAGCAGCAGGAGATTGTGAATCTCTTCTTCATCCCAACCC 599

Qy   1379 AGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCG 1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 AGGCTGTGGGCGCCATCATCGGGAACAAGGGGCACACATCAAACAGCTGGCGAGATTCG 659

Qy   1439 CCGGAGCCCTCTATCAAGATTGCCCCTGCGAAGCCCAGACGTCAGCGAAAGGAATGGTCA 1498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    660  CCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGCTGCAGCCGAAAGATGGTCA  719
                 |||||||:|||| |||||||||||| :||||||||||||||:||||||||:||||||||:
Qy   1499  TCATCACCGGGCCACCGGAA-GCCCAGTTCAAGGCACCGGATCTTTGGGAAACTG  1557
           |||||:|  ||||| ||||:|||||||||||||  ||:||:|| ||||||:|||||||
Db    720  TCATCAMCGGGCCAMCGGAATGCCCAGTTCAAGGCCCAGGACGGATYTTTGGGAAAYTG  779

Qy   1558  AAAG-AGGAAAAACTTCTTTAACCCAAAGAGAAGAAGTGAAGCTG  1599
           ||||  ||||||| :|||||||||||||||||||||||||||||||
Db    780  AAAGAAGAAAAATTTTTTAAACCCAAAGAGAAGAAGTGAAGCTG  822
```

RESULT 15

AAC66035
ID   AAC66035   standard; cDNA; 1740 BP.
XX
AC   AAC66035;
XX
DT   21-FEB-2001   (first entry)
XX
DE   Human lung cancer-associated cDNA antigen L523S.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US008896.
XX
PR   02-APR-1999;   99US-00285479.
PR   17-DEC-1999;   99US-00466396.
PR   30-DEC-1999;   99US-00476496.
PR   10-JAN-2000;   2000US-00480884.
PR   22-FEB-2000;   2000US-00510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T, Fan L;
XX
DR   WPI; 2000-628399/60.
DR   P-PSDB; AAB11365.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient.
XX
PS   Claim 1a; Page 258-259; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer
XX
SQ   Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 U; 0 Other;

Query Match   20.9%; Score 712; DB 3; Length 1740;
Best Local Similarity   64.5%; Pred. No. 3.7e-178;
Matches 1156; Conservative   0; Mismatches 555; Indels   81; Gaps   3;

```
Qy   73  ATGAACAAGCTTTACATCGGAAACTTGAGCCCCGCCGTCACCGCCGACCTCCGGCAG  132
         ||||||||| || |||||||||||| | |||||||||    |||| |||||||||||
Db    1  ATGAACAAACTGTATATCGGAAACCTCAGCGAGAACCTCAGCGAGAACCTCCGGCAG  60
```

```
Qy   133  CTCTTGGGACGAGGAAGCTGCCCCTGCGGGACAGCAGTGTCCTGCTGAAGTCCGGCTACGCC  192
              |||||| || ||||||| ||||:|||||||| || | ||:||||||||||:||||||||:
Db    61  ATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCCTTCCTGCTGAAGACTGGCTACGCG  120

Qy   193  TTCGTGACTACCCGACCAGGACTGGGGCCATCGCGCCTCAAGCACAGAACCGCC  252
          |||||||||| ||||:||||||||||||||||| |||||||  ||||||||||:
Db    121  TTCGTGACTGCCGACGAGCTGGGCGTGGGCCCTCAAGGCCATCGAGCCGCTTCAGTGTAAA  180

Qy   253  GTGGAATTGCATGGAAAATCATGGAAGTTGATTACTCAGTTCTTAAAAAGTAAGGAGC  312
              |||||||||| |||| |||||  ||||||| ||||| ||| || ||||||| |||
Db    181  ATAGAACTGCACGGAAACCCATAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAAGGGATT  240

Qy   313  AGGAAAATTCAGATTCGAAACATCCTCTCACCTGCAGTGGGAGGTGTTGGATGGGACTT  372
          :|||||||||||:||  ||||||||| ||||||  |||||||| ||||||  |||| |||
Db    241  CGGAAACTTCAGATACGCAAAATATCCCCGCCCCCATTTACAGTGGGAGGTGCTCGATAGTTTA  300

Qy   373  TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTGCAACACAGAACAGAACCGCC  432
           | ||||||:|||||:||||||| ||| ||||:||||| |||||||||| |:||||||||
Db    301  CTAGTCCAGTATGGAGTGGTGAGAGCTGTGAGCAAGTGAAGTGAAACTGCA  360

Qy   433  GTTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGAGAAGCTAAGC  492
          ||||| |||| ||| | ||||| ||:|| |||| || |||  ||| | ||||  || |||
Db    361  GTTGTAAATGTAACCTATTCCAGTAGGACCAAGCTAGACAAGCACTAGACAAACTGAAT  420

Qy   493  GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGATGAAGAGGTGAGC  552
          ||| || |||||||||| ||:| ||| || ||||| || |||:|| |||| |||| ||
Db    421  GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAACGGCCGCC---  480

Qy   553  TCCCCTTCGCCCCCCTCAGCGAGCCCCAGCTGCGGGACCAAGGCCACCTCTTCCCGGAGCAAGGC---  609
          ||||   | | |||: ||| || || ||:||  ||||||||||||||||| ||||||||||||
Db    481  CAGCAAACCCCTTGCAGCAGCCCCCAGGTCGCCGGGGCTTGGGCAGAGGGGCCTGGGGGCTCCTCA  540

Qy   610  ------CACGGCCCCTGGGGCCACTTCTCAGGCCACAGATTGATTTCCGTCGCGGATC  663
                |||||||||||| ||||||||||| ||:||||||:| ||||||||:|| |||||
Db    541  AGGCAGGGGTCTTCCAGGATCCGTATCCAAGCAGAGAAACCATGTGATTTGCCTCTGCGCCTG  600

Qy   664  CTGGTCCCCCAGTTGTTGGTGTCGCCATCATCGGAAAAGAGGGCTTGACCATAAAGAAC  723
          ||||:|| ||||| ||| || | |:||||| |||||||| | ||| || ||| |||| |
Db    601  CTGGTTCCCACCCAATTTGTTGGAGCCATCATAGGAAAAGAGAAGGTGCCACCATTCGGAAC  660

Qy   724  ATCACTAAGCAGACACCAGTCCCGGTAGATATCCATAGAAAAGAACTCTGGAGCTGCA  783
          |||| |  | || || |||| ||||   || |||| | ||| |  || ||||| |||
Db    661  ATCACCAAACAGACACCAGTCAGTGTCTAAAATCGATGTCCACCGTAAAGAAAATGCGGGGGCTGCT  720

Qy   784  GAGAAGCCTGTCACCATCCATGCCACCCCAGAGAGGGACTTCTGAAGCCATGCCGCATGATT  843
          |||||| || ||  | ||| ||||||| ||  ||||| |||| ||||||||| | ||:||||
Db    721  GAGAAGTCGATTACTATCTCCTCTACTCCTGAAGGCACCCTCTGCGGCTTGTAAGTCTATT  780

Qy   844  CTTGAAATCATGCAGAAGAGGACAGATGAGCAACAAACTAGCCAGATATAAATTCACAGAGATCCCCTTGAAG  903
          |||||| | |  |||  ||  ||| ||| |  |||  | | || |||| |||:||||||| |  ||||||
Db    781  CTGGAGATTATGCATAAGGAGCTCAAGATTCAAGATATAAAATTCACAGAAGAGATCCCCTTGAAG  840

Qy   904  ATCTTGGCACCACAATGGCTTGGTTGGAAGACTGATTGGAGAAAGGAGGCAGAAATTTGAAG  963
          ||:||  | |||| |  | || ||||||| || || ||:|||| | ||||||||| ||||||
Db    841  ATTTTAGCTCATAATAACTTGTTGGACGTCTTATTGGTAAAGAAGGAGAAAATCTTAAA  900

Qy   964  AAAATTGAACATGAAACATGAACAGGGACCAAGATAACAATCCATCTTTGGAGATTGAGCATA  1023
          |||||||||| | ||  || |||| ||  | |||| ||||  ||| ||:|| |||| |||||
Db    901  AAAATTGAGACAAGAACTAAAATCACGATATCTCCATTGCCATTGCAGGAATTGCACGCTG  960

Qy   1024  TACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCGTGTCGCCAGTGCTGAG  1083
           || | ||| || || || |  ||| | ||||| ||||| | |||  |||||| ||| |||
Db    961  TATAATCCAGAACCGCCACTATTACAGTTAAAGGCAATGTTGAGACATGTGCCAAGCTGAG  1020

Qy   1084  ATAGAGATTATGAAGAAGTGCGTGAGGCCTTGAGGCCTGTGTTAACCAA  1143
           | |||||||||||| |||| | |  |||| |||| |||| | |||| ||
Db    1021  GAGGAGATCATGAAGAAAATGCTTATGCAGGAGTCTTATGGAAAAAT GATATTGCTTCTATGAATCTT  1080

Qy   1144  CAAGCCAATCGATCCCAGGTTGAACCTCAGCCGCATCTTTTCACAGGACTG  1203
           |||| | || || ||||| |   ||| |  | |||| ||:|| |||| |
Db    1081  CAAGCACATTTAATTCCTGGATAAATCTGAACGCCCTTGGTCGTCTTGTTCCCACCCACTTCA  1140

Qy   1204  TCCGTGCTATCTCCACCAGCAGCAGGGCCCCGCGGGAGCTCCCCGCCTGCCCCCTACCACCCC  1263
```

```
Db   1141  GGGATGCCACCTCCCACCTCCAGCGGGGCCCCCTTCAGCCATGACTCCT------------- 1185
                 ||||   |||  ||  ||||||||||||| ||
Qy   1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCGTTC 1323
                 |||-|-| -||||||
Db   1186  ------------------------------------CCCTACCCGCAGTTTGAGC----- 1204

Qy   1324  CCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCT 1383
                 ||--||-||  |||| ||  ||
Db   1205  -------------AATCAGAAACGGAGACTGTTCATCTGTTTATCCCAGCTCTATCA 1248

Qy   1384  GTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGA 1443
                 ||||||||||||||-||||||
Db   1249  GTCGGTGCCATCATCGGCAAGCAGGGCCAGCACATCAAGCAGCTTTCTCGCTTTGCTGGA 1308

Qy   1444  GCCTCTATCAAGAGATTGCCCCTGCGGAAGGGCCCAGACGTTCAGCGGAAAGGATGGTCATCATC 1503
                 |-|||||| |||||| -||||| |-|
Db   1309  GCTTCAATTAAGATTGCTCAGCCGAAGCACCAGATGCTAAAGTGAGGATGGTGATTATC 1368

Qy   1504  ACCGGGCCACCGGAAGCCCAGTTCAGGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAG 1563
                 ||| |||||||||||-|||||
Db   1369  ACTGGACCACCAGAGGCTCAGTTCAGGGCTTCAGGGAAGAATTTATGGAAAAAATTAAAGAA 1428

Qy   1564  GAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
                 ||||| |||||||| -|||||
Db   1429  GAAAACTTTGTTAGTCCTAAAGAAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCCATCC 1488

Qy   1624  TCCACAGCTGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTA 1683
                 ||||||| -||||||
Db   1489  TTTGCTGCTGGCAGAGTTATTGGAAAAAGGAGGCAAAAACGGTGAATGAACTTCAGAATTTG 1548

Qy   1684  ACCAGTGCAGAAGTCATCGTGCCTCGTGCACCAAAAGCCCAGATGAAAAATGAGGAAGTGATC 1743
                 ||||||||||| ||| ||||||
Db   1549  TCAAGTGCAGAAGTTGTGTCCCTCGTGACCAGACACCTGATGAGAATGACCAAGTGGTT 1608

Qy   1744  GTCAGAATTATCGGGCACTCTTCTTTGCTAGCCAGACTGCACAGGCGCAAGATCAGGGAAATT 1803
                 ||-|||| -|||||
Db   1609  GTCAAAATAACTGGTCACTTCTATGCTTGCCAGGTTGCCCAGAGAAAAATTTCAGGAAATT 1668

Qy   1804  GTACAACAGGTGAAGCAGCAGCAGGAGCAGGAAAATACCCTCAGGGAGTCGCCTCAC 1855
                 |||||| |||||||
Db   1669  CTGACTCAGGTAAAGCAGCAAGCACCAACAACAGAAGGGCTCTGCAAAGTGGACCAC 1720
```

Search completed: July 21, 2004, 00:08:59
Job time : 1254 secs

```
OM nucleic - nucleic search, using sw model

          Copyright (c) 1993 - 2004 Compugen Ltd.
                    GenCore version 5.1.6

Run on:    July 20, 2004, 23:48:10 ; Search time 226 Seconds
                              (without alignments)
                              8378.285 Million cell updates/sec

Title:         US-09-270-437D-6
Perfect score: 3412
Sequence:      1 ggcagcggaggaggcgagga........aaccttgaaatgtttatt 3412

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched:      682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1355418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued_Patents_NA:*
              1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
              2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
              3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
              4: /cgn2_6/ptodata/2/ina/6E_COMB.seq:*
              5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	3411	100.0	3412	3	US-09-061-709-6	Sequence 6, Appli
2	3411	100.0	3412	4	US-09-899-651-6	Sequence 6, Appli
3	3143	92.1	3283	3	US-09-061-709-8	Sequence 8, Appli
4	3143	92.1	3283	4	US-09-899-651-8	Sequence 8, Appli
5	712	20.9	1740	4	US-09-643-597-347	Sequence 347, App
6	712	20.9	1740	4	US-09-542-615A-347	Sequence 347, App
7	712	20.9	1740	4	US-09-606-421B-347	Sequence 347, App
8	710.4	20.8	4159	3	US-09-061-709-4	Sequence 4, Appli
9	710.4	20.8	4159	4	US-09-899-651-4	Sequence 4, Appli
10	710.4	20.8	4181	4	US-09-643-597-175	Sequence 175, App
11	710.4	20.8	4181	4	US-09-480-884A-175	Sequence 175, App
12	710.4	20.8	4181	4	US-09-542-615A-175	Sequence 175, App
13	710.4	20.8	4181	4	US-09-606-421B-175	Sequence 175, App
14	691.8	20.3	2224	4	US-09-261-855-1	Sequence 1, Appli
15	570.6	16.7	1946	3	US-09-061-709-7	Sequence 7, Appli
16	570.6	16.7	1946	4	US-09-899-651-7	Sequence 7, Appli
17	489.2	14.3	1708	3	US-09-061-709-5	Sequence 5, Appli
18	489.2	14.3	1708	4	US-09-899-651-5	Sequence 5, Appli
19	223.8	6.6	317	4	US-09-621-976-13846	Sequence 13846, A
20	73.8	2.2	7218	1	US-08-232-463-14	Sequence 14, Appl
21	56.8	1.7	7218	4	US-08-232-463-14	Sequence 14, Appl
22	45	1.3	598	4	US-09-669-751-39	Sequence 39, Appli
23	44.6	1.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
24	44.6	1.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
25	44.4	1.3	5562	4	US-10-204-708-63	Sequence 63, Appli
c 26	44	1.3	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 27	43.4	1.3	55298	4	US-09-491-356C-1	Sequence 1, Appli
c 28	42.8	1.3	289	3	US-09-007-005-17	Sequence 17, Appl
c 29	42.8	1.3	289	3	US-09-244-796-17	Sequence 17, Appl
c 30	42.8	1.3	17056	3	US-09-245-041-3	Sequence 3, Appli
c 31	42.2	1.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
c 32	42.2	1.2	80161	3	US-09-370-700-1	Sequence 1, Appli
c 33	42.2	1.2	80161	4	US-09-603-207-1	Sequence 1, Appli
c 34	42	1.2	277	3	US-09-007-005-3	Sequence 3, Appli
c 35	42	1.2	277	3	US-09-244-796-3	Sequence 3, Appli
c 36	42	1.2	832	4	US-09-621-976-2813	Sequence 2813, Ap
c 37	41.6	1.2	248	3	US-09-007-005-32	Sequence 32, Appl
c 38	41.6	1.2	248	3	US-09-244-796-32	Sequence 32, Appl
c 39	41.4	1.2	791	3	US-08-998-416-346	Sequence 346, App
40	41.4	1.2	7130	3	US-09-056-105-31	Sequence 31, Appl
41	41.2	1.2	362	3	US-09-018-584A-11	Sequence 11, Ap
42	41.2	1.2	1722	4	US-09-134-001C-1976	Sequence 1976, Ap
43	41.2	1.2	56520	4	US-09-338-907-179	Sequence 179, APP
44	41.2	1.2	56520	4	US-09-218-207-179	Sequence 179, APP
45	41	1.2	3138	1	US-07-867-106-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
;  TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/061,709B
;  CURRENT FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
SEQ ID NO 6
;  LENGTH: 3412
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-061-709-6

  Query Match             100.0%;  Score 3411;  DB 3;  Length 3412;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCCGCGGGGAGCGGCGCGGGCTCGGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGCCGCGGGGAGCGGCGCGGGCTCGGGG 60

Qy     61  AAGAGACGGATGATGAACAAGCTTTACATCGGGACAAGGAACCTGAGCCCCGCCGCCGAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  AAGAGACGGATGATGAACAAGCTTTACATCGGGACAAGGAACCTGAGCCCCGCCGCCGAC 120

Qy    121  GACCTTCCGGCAGCTCTTTGGGACACAGGAAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GACCTTCCGGCAGCTCTTTGGGACACAGGAAGCTGCCCCTGCGGGACAGGTCCTGCTGAAG 180

Qy    181  TCCGGCTACGGCCTTCGTGGACTACCCGACCAGAACTCGGCCATCGAGACTCCGAGACC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TCCGGCTACGGCCTTCGTGGACTACCCGACCAGAACTCGGCCATCGAGACTCCGAGACC 240

Qy    241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTTCAGTCTTAA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTTCAGTCTTAA 300

Qy    301  AAGCTTAAGGAGCAGGAGAAATTCAGATTCAGAATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
```

```
Db

Qy   301 AAGCTAAGGAGGACGCAGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGGAGGTG 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Qy   361 TTGGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGACAAGTCAACACAGAC 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGACAAGTCAACACAGAC 420

Qy   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGCAAAAATAGCCATG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGCAAAAATAGCCATG 480

Qy   481 GAGAAGCTAAGGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGAT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGCTAAGCGGGCATCAGTTGAGAACTCAGTTCCTTCAAGATTCCTACATCCCGGAT 540

Qy   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCGG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTGCGCCCCTCAGCGAGCCCAGCGTGGGACCACTCTTCCCGG 600

Qy   601 GAGCCAAGGCCACGCCCCGTGGGGCACTTCTCAGCGGCCAGACAGATTGATTCCCGTGCG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCCAAGGCCACGCCCCGTGGGGCACTTCTCAGCGGCCAGACAGATTGATTCCCGTGCG 660

Qy   661 ATCCTGTCCCCAGTTTGTTGGTGCCATCATGGAAAGGAGGCTTGACCATAAAG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGTCCCCAGTTTGTTGGTGCCATCATGGAAAGGAGGCTTGACCATAAAG 720

Qy   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGTATATCCATAGAAAAGAACTCTGGAGCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCAGACCCAGTCCCGGGTAGTATATCCATAGAAAAGAACTCTGGAGCT 780

Qy   781 GCAGAGAAGCCTGTCACCATCCATCATCGCGACTTCTGAAGCATGCCGCATG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATCCATCATCGCGACTTCTGAAGCATGCCGCATG 840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy   901 AAAATCTTGGCCACACAATGGCTTGGTTGGTGGAAGACTGATTGGAAAAGAAGGCAGAAATTG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGGCCACACAATGGCTTGGTTGGTGGAAGACTGATTGGAAAAGAAGGCAGAAATTG 960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGACCAAGAAGATACAACAGTATACAAATCTCATCTTTGCGAGGATTTGAGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGACCAAGAAGATACAACAATCTCATCTTTGCGAGGATTTGAGC 1020

Qy  1021 ATATACAACCCGGAAGAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCCGGAAGAGAACCATCACTGTGGGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT 1080

Qy  1081 GAGATAGAGATTATGAAGAAGACTGCGTGAGGCCCTTTGAAAATGATATGCTGGCTGTGTTAAC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAAGACTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTGTTAAC 1140

Qy  1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTTTCAACAGGA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAACAAGCCAATCTGATCCCAGGTTGAACCTCAGCGCACTTTTCAACAGGA 1200

Qy  1201 CTGTCCGTGCTATCTCCACCACTCCGATATCTTCCTGTACCCCGCGTCCCCCGTACCAC 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CTGTCCGTGCTATCTCCACCACTCCGATATCTTCCTGTACCCCCGCGTCCCCCGTACCAC 1260

Qy  1261 CCCTTCACTACCCACTCCGATACCTCTTATCGGCAGGAGATTGTGAATCTCCCCAACCCAGTTTGGCCCCG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCCTTCACTACCCACTCCGATACCTCTTATCGGCAGGAGATTGTGAATCTCCCCAACCCAGTTTGGCCCCG 1320

Qy  1321 TTTCTTTTCTTTTAAAAGAAGAAAGCAGGCTTTGAATCTTCTCCTTCATCCCCAACCCAG 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TTTCTTTTCTTTTAAAAGAAGAAAGCAGGCTTTGAATCTTCTTCATCCCCAACCCAG 1380

Qy  1381 GCTGTGGGCGGCCCATCATCGGGAAGAAGAAGGGGCACACAATCAAAACTGGCCGAGATTCGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1381 GCTGTGGGCGGCCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCCGAGATTCGCC 1440

Qy  1441 GGAGCCTCATCAAGATGCCCCTGCGCCCCGAAGGCCCAGCCCCAGACGTCAGCGAAAGGATGGTGCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GGAGCCTCATCAAGATGCCCCTGCGCCCCGAAGGCCCAGCCCCAGACGTCAGCGAAAGGATGGTCATC 1500

Qy  1501 ATCACCGGCCACCGGAGCCGAGCCAGTTCAAGGCCCAGGCGACGGATCTTTGGGAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ATCACCGGGCCACCGGAGCCGAGCCAGTTCAAGGCCCAGGCGACGGATCTTTGGGAAACTGAAA 1560

Qy  1561 GAGGAAAACTTCTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAAGTGCCC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCATATCAGAAGTGCCC 1620

Qy  1621 TCTTCCACAGTCGGCCGGTGATTGGCAAGAACGTGGCAAGACCGTGAACGAACTGCAGAAC 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TCTTCCACAGTCGGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAAC 1680

Qy  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGAGCAGCCAGAGTCGCACACGGGAA 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGAGCAGCCAGAGTCGCACACGGGAA 1740

Qy  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCAGCGAGGAGTCGCCTCACAGCGC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTAGCAGCGAGGAGTCGCCTCACAGCGC 1800

Qy  1801 ATTGTACAACAGGTGAAGCAGCAGAGGCACCAGCAGCACCAGGAGTCGCCTCACAGCGC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATTGTACAACAGGTGAAGCAGCAGAGGCACCAGCAGCACCAGGAGTCGCCTCACAGCGC 1860

Qy  1861 AGCAAGTGAGGCTCCCACAGCAACCAGCAAACAACGGATGAATGTAGCCCTTCCAACAC 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AGCAAGTGAGGCTCCCACAGCAACCAGCAAACAACCAGATGAATGTAGCCCTTCCAACAC 1920

Qy  1921 CTGACAGAATGAGACCAAACGAGCAGATCGCGGGAGCAAACCAAAGACCATCTGAG 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGACAGAATGAGACCAAACGAGCAGATCGCGGGAGCAAACCAAAGACCATCTGAG 1980

Qy  1981 GAATGAGAAGTCTGCGGAGGCGGCCCCAGGGACTCGCGAGGGCCTGAGACCCCAGGGGC 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GAATGAGAAGTCTGCGGAGGCGGCCCCAGGGACTCGCGAGGGCCTGAGACCCCAGGGGC 2040

Qy  2041 CGAGGAGGGCGGGGAAGGTCAGCCAGTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 CGAGGAGGGGCGGGGAAGGTCAGCCAGTTGCCAGAACCACCGAGCCCCGCCTCCCGCC 2100

Qy  2101 CCCAGGGCTTCTGCAGCTTGAGGCCTGTGCCAGTGCTCCTCGGATCTCTCCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 CCCCAGGGCTTCTGCAGCTTCAGGCCTGTGCCAGTGCTCCTCGGATCTCTCCTGAA 2160

Qy  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CTCCCAGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy  2221 AAAATGCACACCCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATTATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AAAATGCACACCCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATTATTAGAAA 2280

Qy  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCCTGTGGGTTACACAGGGTAATATAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCCTGTGGGTTACACAGGGTAATATAT 2340

Qy  2341 TTTAGAAATAATAATATCAAAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATAATATCAAATAATCAACTAACTCCAATTTTAATCAATTATTAATTTT 2400

Qy  2401 TTTTTCTTTTAAAAGAAGAAAGCAGGCTTTGAGTCGAGACTTTAAAGAATAAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTTCTTTTAAAAGAAGAAAGCAGGCTTTGAGTCGAGACTTTAAAGAATAAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTAGAGGAGCTGTTTGAGGCTGAGGAGCTGTTTGAGGCGCACAAAATTCACCCAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTAGAGGAGCTGTTTGAGGCGCACCCGCACAAAATTCACCCAGAGGGAAAT 2520
```

```
Query Match             100.0%;  Score 3411;  DB 4;  Length 3412;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 GGCAGCGGAGGAGGCGGAGGAGCGGAGGAGGCCGGGGAGCCGGGGCCGGGGCTTCGGGGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCGGAGGAGGCGGAGGAGCGGAGGAGGCCGGGGAGCCGGGGCCGGGGCTTCGGGGG 60

Qy   61 AAGAGACCGATGATGAACAAGCTTACATCGGGAGAGTCGCCCTGCCCGGGACAGTCCTGCTGAAG 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACCGATGATGAACAAGCTTACATCGGGAGAGTCGCCCTGCCCGGGACAGTCCTGCTGAAG 120

Qy  121 GACCTCCGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGTCCGTGCTGAAG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GACCTCCGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGTCCGTGCTGAAG 180

Qy  181 TCCGGCTACGCCCTTCGTGGACTACCCCGACCCCGAACTGGGCCATCCGCGCCATCGAGACC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCCTTCGTGGACTACCCCGACCCCGAACTGGGCCATCCGCGCCATCGAGACC 240

Qy  241 CTCTCGGTAAAGTGGAATTGCATGGGAAAAATCATGGAAAATCATGAAGTCAGTCTCTAAA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGTAAAGTGGAATTGCATGGGAAAAATCATGGAAAATCATGAAGTCAGTCTCTAAA 300

Qy  301 AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCCTCCAGTGAGTGGGAGGTG 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGCTAAGGAGCAGGAGAAATTCAGATTCGAAACATCCCCTCCAGTGAGTGGGAGGTG 360

Qy  361 TTGATGGACTTTTGGCTCAATATGGGACAGTCAGTTGAGAATGTGGAACAGTCAACACAGAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGATGGACTTTTGGCTCAATATGGGACAGTCAGTTGAGAATGTGGAACAGTCAACACAGAC 420

Qy  421 ACAGAAACCGCCGTGTCAAGTGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATG 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ACAGAAACCGCCGTGTCAAGTGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATG 480

Qy  481 GAGAAGTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCTACATCCGGAT 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGAAGTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCTACATCCGGAT 540

Qy  541 GAAGAGGTGAGCTCCCCTCGCCCGAGCCCAGCGTGGGACCCACTCTTCCCGG 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAAGAGGTGAGCTCCCCTCGCCCGAGCCCAGCGTGGGACCCACTCTTCCCGG 600

Qy  601 GAGCAAGGCCACGCCCGCTGGGGGCACTTCTCAGGCCAGCACAGATTGATTCCCGCTGCGG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GAGCAAGGCCACGCCCGCTGGGGGCACTTCTCAGGCCAGCACAGATTGATTCCCGCTGCGG 660

Qy  661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGCTTGACCATAAAG 720
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGCTTGACCATAAAG 720

Qy  721 AACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACATCACTAAGCAGACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
```

```
Qy 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATA 2580
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAGAGGGATA 2580

Qy 2581 CCGTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCTCAGCTCATACACCCATT 2640
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 CCGTCTCCTGAAGAGGAAACTCTGTCACTCCTCAGCCTGTCTCAGCTCATACACCCATT 2640

Qy 2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCCTCTGTCT 2700
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 TCTCTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATACTGCTATATAATTCCTCTGTCT 2700

Qy 2701 CTCTTGTTTATCTCTCCCCTCCCCCCTCCCTTCTTCTCCATCCCATTCTTTTGAA 2760
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 CTCTTGTTTATCTCTCCCCTCCCCCCTCCCTTCTTCTCCATCCCATTCTTTTGAA 2760

Qy 2761 TTTCCTCATCCTCCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCAGGCAAAGCA 2820
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 TTTCCTCATCCTCCCATCTCAATCCCGTATCTACGGCACCCCCCCCCCCCCAGGCAAAGCA 2820

Qy 2821 GTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAACCAGCCTCAA 2880
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 GTGCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAACACACAACCAGCCTCAA 2880

Qy 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG 2940

Qy 2941 AGGAAAAACAGGAACCCACCAAACCACCAATCAACCAAACAAAGAAAAAATTCCACAATG 3000
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 AGGAAAAACAGGAACCCACCAAACCACCAATCAACCAAACAAAGAAAAAATTCCACAATG 3000

Qy 3001 AAAGAAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 AAAGAAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy 3061 ATTCCTTTCTTAAAAAAAAAATGTGGAGGAAAGTAGAATTACCAAGGTTGTTGGCC 3120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 ATTCCTTTCTTAAAAAAAAAATGTGGAGGAAAGTAGAATTACCAAGGTTGTTGGCC 3120

Qy 3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAG 3180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAG 3180

Qy 3181 GTGTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTGATAG 3240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3181 GTGTTTTACCTCAGCACCTTGCTCTTGTGTTCCCTTAGAGATTTTGTAAAGTGATAG 3240

Qy 3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAGATATCAACT 3300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3241 TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAGATATCAACT 3300

Qy 3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGCAACAGCTGTTCTTCCGCT 3360
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 3301 GCCAGCCTGGAGAAGTGACAGTCCAAGTGCAACAGCTGTTCTTCCGCT 3360

Qy 3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3361 AGCCAAGAACNATATGGCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT 3412
```

RESULT 2
US-09-899-651-6

```
Qy   781  GCAGGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GCAGGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATG  840

Qy   841  ATTCTTGAAATCATGCAGAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGAATTCCTCTG  900
          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   841  ATTCTTGAAATCATGCAGAAGAGAGCAGATGAGACCAAACTAGCCGAAGAGAATTCCTCTG  900

Qy   901  AAAATCTTGGCACACAAATGGCTTGGTTGGAAGAAGAAGGCAGAAATTTG  960
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  AAAATCTTGGCACACAAATGGCTTGGTTGGAAGAAGAAGGCAGAAATTTG  960

Qy   961  AAGAAAATTGAACATGAAACAGGGACCAAGTAACAATCTCATCTTTGCAGGATTTGAGC  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AAGAAAATTGAACATGAAACAGGGACCAAGTAACAATCATCTTTGCAGGATTTGAGC  1020

Qy  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGCCAGTGCT  1080

Qy  1081  GAGATAGAGATTATGAAGAAGCTGCGTAGGCCTTTGAAAATGATATGTGGCTGTGTTAAC  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GAGATAGAGATTATGAAGAAGCTGCGTAGGCCTTTGAAAATGATATGTGGCTGTGTTAAC  1140

Qy  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGA  1200

Qy  1201  CTGTCCGTGCTATCTCCACCACTCCGGATAGTTCTTCCAGCAGGAGGGGCACACTACCAC  1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CTGTCCGTGCTATCTCCACCACTCCGGATAGTTCTTCCAGCAGGAGGGGCACACTACCAC  1260

Qy  1261  CCCTTCACTACCACCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCG  1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  CCCTTCACTACCACCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCG  1320

Qy  1321  TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCAACCCAG  1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  TTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCAACCCAG  1380

Qy  1381  GCTGTGTGGGGGCCATCATCGGGGAAGAAGGGGGCACACATCAAACAGCTCGGCGAGATTGGCC  1440
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  GCTGTGTGGGGGCCATCATCGGGGAAGAAGGGGGCACACATCAAACAGCTCGGCGAGATTGCCC  1440

Qy  1441  GGAGGCCTCTATCAAGATTGCCCCGGAAGCCCAGAGCGTCAGCGGAAGGATGGTCATC  1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  GGAGGCCTCTATCAAGATTGCCCCGGAAGCCCAGAGCGTCAGCGGAAGGATGGTCATC  1500

Qy  1501  ATCACCGGGCACCCGAAGCTTCAAGGCCCAGTTCAAGGCGGATCTTTGGGAAACTGAAA  1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  ATCACCGGGCACCCGAAGCTTCAAGGCCCAGTTCAAGGCGGATCTTTGGGAAACTGAAA  1560

Qy  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCTGAAGCGCATATCGAGTGCCC  1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGGAAGCTGAAGCGCATATCGAGTGCCC  1620

Qy  1621  TCTTCCACAGCTGCCGGGGTGATTGGCCAAAGGTGGCAAGACCGTGCAAGAAC  1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  TCTTCCACAGCTGCCGGGGTGATTGGCCAAAGGTGGCAAGACCGTGCAAGAAC  1680

Qy  1681  TTAACCAGTGCAGAAGTGCAGAAGTCAGTTCTTTGCTAGCGAGCAGAAATACCCTGACCAAAATGAGGAAGTG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  TTAACCAGTGCAGAAGTGCAGAAGTCAGTTCTTTGCTAGCGAGCAGAAATACCCTGACCAAAATGAGGAAGTG  1740

Qy  1741  ATCGTCAGAATTATCGGGCACTCTTTTGCTAGCGAGCAGACTGCACAGGCGCCAAGAGTCGGCAAGATCAGGGAA  1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  ATCGTCAGAATTATCGGGCACTCTTTTGCTAGCGAGCAGACTGCACAGGCGCCAAGAGTCGGCAAGATCAGGGAA  1800

Qy  1801  ATTGTACAACAGGTTGAAGCAGCAGGAGCAGGAGCAGGGAGTCGCCTTCACAGCGC  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  ATTGTACAACAGGTTGAAGCAGCAGGAGCAGGAGCAGGGAGTCGCCTTCACAGCGC  1860

Qy  1861  AGCAAGTGAGGCTCCCAGGCACCAGCAAACAACGATGAATGTAGCCCTTCCAACAC  1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  AGCAAGTGAGGCTCCCAGGCACCAGCAAACAACGATGAATGTAGCCCTTCCAACAC  1920

Qy  1921  CTGACAGAATGAGACCAAACGAGCCAGCAGCCAGCCAGTGGGAGCAAACCAAAGACCATCTTGAG  1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  CTGACAGAATGAGACCAAACGAGCCAGCAGCCAGCCAGTGGGAGCAAACCAAAGACCATCTGAG  1980

Qy  1981  GAATGAGAGTCTGCGAGGCGGCCCAGGGACTCTGCCGAGGCCTGAGAACCCCAGGGGC  2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  GAATGAGAGTCTGCGAGGCGGCCCAGGGACTCTGCCGAGGCCTGAGAACCCCAGGGGC  2040

Qy  2041  CGGAGGAGGGCGGGGGAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGGCTCCCGCC  2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  CGGAGGAGGGCGGGGGAGGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGGCTCCCGCC  2100

Qy  2101  CCCCAGGGCCTTCTGCAGCCATCCAGCTTCACTTCACCATCCACTCACTCCTGAA  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CCCCAGGGCCTTCTGCAGCCATCCAGCTTCACTTCACCATCCACTCACTCCTGAA  2160

Qy  2161  CTCCCACAGCGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  CTCCCACAGCGTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220

Qy  2221  AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2280

Qy  2281  GGGAAGAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATAT  2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  GGGAAGAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATAT  2340

Qy  2341  TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT  2400
          |||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TTTAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTT  2400

Qy  2401  TTTTCTTTTTAAAGAAAGCAGCTTTTAAAGAATAAAGTCTTTGGGAG  2460
          |||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TTTTCTTTTTAAAGAAAGCAGCTTTTAAAGAATAAAGTCTTTGGGAG  2460

Qy  2461  GTCTCACGGTGTAGAGGACACTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGGGAAAT  2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GTCTCACGGTGTAGAGGAGAGCTTCACGGCAGTTCTGATCACCGTGTATGTCAACAGAGGGAAAT  2520

Qy  2521  CTCGTCCGAAGGACACTCACGGCAGTTCTGGATCACCTGTCGCTGTCTAGCTCATCATCACCCATT  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CTCGTCCGAAGGACACTCACGGCAGTTCTGGATCACCTGTCGCTGTCTAGCTCATCATCACCCATT  2580

Qy  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACCTCCTACTCGTCTGTTTTTGCATACTGGTATATAATTCTCTGTCT  2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACCTCCTACTCGTCTGTTTTTGCATACTGGTATATAATTCTCTGTCT  2640

Qy  2641  TCTCTTTGCTCACAGGTTTTAAACTGGTTTTTGCATACTGCTATATATAATTCTCTGTCT  2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  TCTCTTTGCTCACAGGTTTTCACAGGTTTTTGCATACTGCTATATATAATTCTCTGTCT  2700

Qy  2701  CTCTCGTTTATCTTCCATCTCCATCTCCCCTTCTTCTTCTTCCATTCTTTTGAA  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  CTCTCGTTTATCTTCCATCTCCATCTCCCCTTCTTCTTCTTCCATTCTTTTGAA  2760

Qy  2761  TTTCCTACCCTCCAATCCGTATCTCACGGCACCCCCCCCCCCCAGGCAAAGCA  2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  TTTCCTACCCTCCAATCCGTATCTCACGGCACCCCCCCCCCCCAGGCAAAGCA  2820

Qy  2821  GTTGCTCGAGTATCACATCACACAAAGGAACAAAGAGAAACCAGCTCAA  2880
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  GTTGCTCTGAGTATCACATCACACAAAGGAACAAAGAGAAACCAGCTCAA  2880

Qy  2881  CTTACACTTGGTTACTCAAAGATCAATGGTACTTGTCCTAGCGTTTGGAAG  2940
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  CTTACACTTGGTTACTCAAAGATCAATGGTACTTGTCCTAGCGTTTGGAAG  2940

Qy  2941  AGGAAAACAGGAACCCACCAAACCAATCAACCAACAAGAAAAATTCCACAATG  3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2941  AGGAAAACAGGAGACCCACCAAAACCAACCAATCAACCAAACAAGAGAAAAATTCCACAATG  3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181  TCCGGCTACGCCTTCGTGGACTACCCGGCAACCAGAACTGGGCCATCCGCGCCATCGAGACC  240

Qy  3001  AAAGAATGTATTTTGTCTTTTTGTCATTTTGGTGTGTATAAGCCATCAATATTCAGCAAAATG  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300

Db  3001  AAAGAATGTATTTTGTCTTTTTGTCATTTTGGTGTGTATAAGCCATCAATATTCAGCAAAATG  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   241  CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA  300

Qy  3061  ATTCCTTTCTTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  AAGCTAAGGAGCAGGAGAAAATCAGATTCAGAAACATCCTCCTCCACCTGCAGTGGGACGTG  360

Db  3061  ATTCCTTTCTTTAAAAAAAAAAATGTGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301  AAGCTAAGGAGCAGGAGAAAATCAGATTCAGAAACATCCTCCTCCACCTGCAGTGGGAGGTG  360

Qy  3121  CAGGGCGGTTAAATTCACAGATTTTTTAACGAGAGAAAAACACACAGAGAAGAAGCTACCTCAG  3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  TTGGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGGAACAAGTCAACACAGAC  420

Db  3121  CAGGGCGGTTAAATTCACAGATTTTTTAACGAGAGAAAAACACACAGAGAAGAAGCTACCTCAG  3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361  TTGGATGGACTTTTGGCTCAATATGGGACAGTGGACAGTGGAACAAGTCAACACAGAC  420

Qy  3181  GTGTTTTTTACCTTGCAGCACCTTGCTCTTGTGTTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGAAGCAAAAATAGCCATG  480

Db  3181  GTGTTTTTTACCTTGCAGCACCTTGCTCTTGTGTTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421  ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAAGAAGAAGCAAAAATAGCCATG  480

Qy  3241  TTGGAGACATTTTTTTATTTTTAATAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GAGAAGCTAAGGCGGGCATCAGTTGAGAACTACTCCTTGAGCCCAGCCACTCTTCCCGG  540

Db  3241  TTGGAGACATTTTTTTATTTTTAATAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   481  GAGAAGCTAAGGCGGGCATCAGTTGAGAACTACTCCTTGAGCCCAGCCACTCTTCCCGG  540

Qy  3301  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGGCT  3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGACCACCTCTTCCCGG  600

Db  3301  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGGCT  3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   541  GAAGAGGTGAGCTCCCCTTCGCCCCTCAGCGAGCCCAGCGTGGGACCACCTCTTCCCGG  600

Qy  3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
          |||||||||||||||||||||||||||||||||
Db   601  GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTCCCGCTGCGG  660

Db  3361  AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
          |||||||||||||||||||||||||||||||||
Qy   601  GAGCAAGGCCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTCCCGCTGCGG  660

                                                          Qy   661  ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG  720
                                                          Db   661  ATCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAG  720

                                                          Qy   721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACCTGGAGCT  780
                                                          Db   721  AACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAGAGAACCTGGAGCT  780

                                                          Qy   781  GCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGAGGGACTTCTGAGCATGCCGCATG  840
                                                          Db   781  GCAGAGAAGCCTGTCACCATCCATGCACCCCAGAGAGGGACTTCTGAGCATGCCGCATG  840

                                                          Qy   841  ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCTCTG  900
                                                          Db   841  ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGACAAACTAGCCGAAGAGATTCCTCTG  900

                                                          Qy   901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAGAGGCAGAAATTTG  960
                                                          Db   901  AAAATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGAGAGGCAGAAATTTG  960

                                                          Qy   961  AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020
                                                          Db   961  AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020

                                                          Qy  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCACAGTTGCCAGTGCT  1080
                                                          Db  1021  ATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAAGGGCACAGTTGCCAGTGCT  1080

                                                          Qy  1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGCGTGAGGCCTTTGAAAATGATATGCTCGGCTGTTAAC  1140
                                                          Db  1081  GAGATAGAGATTATGAAGAAGCTGCGTGAGCGTGAGGCCTTTGAAAATGATATGCTCGGCTGTTAAC  1140

                                                          Qy  1141  CAACAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGA  1200
                                                          Db  1141  --------------------------------------------------------------  1141

                                                          Qy  1201  CTGTCCGTGCTATCTCCACCAGCAGGGCCCCCGCGAGCTCCCCCGCCGAGCTCCCCCGTACCAC  1260
                                                          Db  1141  --------------------------------------------------------------  1141

                                                          Qy  1261  CCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGCCCCG  1320
```

```
Query Match         92.1%;      Score 3143;    DB 3;    Length 3283;
Best Local Similarity 96.2%;    Pred. No. 0;
Matches 3283;    Conservative    0;    Mismatches    0;    Indels  129;    Gaps    1;

Qy     1  GGCAGCGGAGGAGCGGAGGAGCGGAGGAGCGGCGGGTACCGGGCGCCGGGGGAGCCCGGGGCTCCTCCGGGG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GGCAGCGGAGGAGCGGAGGAGCGGAGGAGCGGCGGGTACCGGGCGCCGGGGGAGCCCGGGGCTCCTCCGGGG  60

Qy    61  AAGAGACGGATGATGAACAAGCTTTACATCGGGAACTTGCCCCTGAGACCCCGTCACCGCCGAC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  AAGAGACGGATGATGAACAAGCTTTACATCGGGAACTTGCCCCTGAGACCCCGTCACCGCCGAC  120

Qy   121  GACCTCCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGGCGGGACAGGTCCTGTGAAG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GACCTCCGGCAGCTCTTTGGGGACAGGAAGTGCCCCTGGCGGGACAGGTCCTGTGAAG  180

Qy   181  TCCGGCTACGCCTTCGTGGACTACCCGGCAACCAGAACTGGGCCATCCGCGCCATCGAGACC  240
```

```
Db  1141 ----------ACCCACTTCGGGATACTTCTCCAGCTGTGTACCCCCCATCACCAGTTTGGCCCG 1191

Qy  1321 TTCCCGCCATCATCACTCTTATCGGGAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192 TTCCCGCCATCATCACTCTTATCGGGAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAG 1251

Qy  1381 GCTGTGTGGCGCCATCATCGGGGAAGCGCACACATCAAACAGCTGGCGAGATTCGCC 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252 GCTGTGTGGCGCCATCATCGGGGAAGCGCACACATCAAACAGCTGGCGAGATTCGCC 1311

Qy  1441 GGAGCCTCTATCAAGATTGCCCTCGCGGAAGGCCCAGACGTTCACGGCGAAAGGATCGTCATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312 GGAGCCTCTATCAAGATTGCCCCTCGCGGAAGGCCCAGACGTTCACGGCGAAAGGATCGTCATC 1371

Qy  1501 ATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAACTGAAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372 ATCACCGGGCACCGGAAGCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAAACTGAAA 1431

Qy  1561 GAGGAAAACTTCTTTAACCCCAAAGAGAGAGTGAAGCTGAAGCTCTGAGAGTCAGAGTGCCC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432 GAGGAAAACTTCTTTAACCCCAAAGAGAGAGTGAAGCTGAAGCTCTGAGAGTCAGAGTGCCC 1491

Qy  1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGCAGAAC 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492 TCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGCAAGACCGTGAACGAACTGCAGAAC 1551

Qy  1681 TTAACCAGTGCAGAAGTCATCGTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552 TTAACCAGTGCAGAAGTCATCGTCGTGACCAAACGCCAGATGAAAATGAGGAAGTG 1611

Qy  1741 ATCGTCAGAATTATCGGGCACTCCTTGCTAGCCGAGACAGAGAATATACCCTCAAGATCAGGGAA 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1612 ATCGTCAGAATTATCGGGCACTCCTTCTTGCTAGCCGAGACAGAGAATATACCCTCAAGATCAGGGAA 1671

Qy  1801 ATTGTACACACAGGTGAAGCAGCAGCAGGCCTCCACAGACAGAAAACAACAACCACCAGCGC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1672 ATTGTACACACAGGTGAAGCAGCAGCAGGCCTCCACAGACAGAAAACAACAACCACCAGCGC 1731

Qy  1861 AGCAAGTGAGGCTCCCACAGACCAGCCAAAACGCAGCCAGATGAATGTAGCCCTTCCAACAC 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1732 AGCAAGTGAGGCTCCCACAGACCAGCCAAAACGCAGCCAGATGAATGTAGCCCTTCCAACAC 1791

Qy  1921 CTGACAGAATGAGACAAAACGCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAG 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1792 CTGACAGAATGAGACAAAACGCAGCCAGCCAGATCGGAGCAAACCAAAGACCATCTGAG 1851

Qy  1981 GAATGAGAAGTCTGCGAGGGACTCTGCCGGAGGCCCTGAGAACCCCCAGGGGC 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1852 GAATGAGAAGTCTGCGAGGGACTCTGCCGGAGGCCCTGAGAACCCCCAGGGGC 1911

Qy  2041 CGAGGAGGGCGGGGAAGGTCAGCGAGGTTGCCAGAGTCAGCCAGCCACCACGGAGCCCCGCC 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1912 CGAGGAGGGCGGGGAAGGTCAGCGAGGTTGCCAGAGTCAGCCAGCCACCACGGAGCCCCGCC 1971

Qy  2101 CCCCAGGGCTTCAGCCAGCTTCAGCGCTTCAGCCATTCCACTTCACCATCCGGATCTCTCCTGAA 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1972 CCCCAGGGCTTCAGCCAGCTTCAGCGCTTCAGCCATTCCACTTCACCATCCGGATCTCTCCTGAA 2031

Qy  2161 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2032 CTCCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC 2091

Qy  2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTTACACGTGTGTACATATTAGAAA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2092 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTTACACGTGTGTACATATTAGAAA 2151

Qy  2281 GGGAAGATGTTAAGATATGTTAAGATATATCAAATAACTAACTCCAATTTTTAATCAATTATTAATTTT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2152 GGGAAGATGTTAAGATATGTTAAGATATATCAAATAACTAACTCCAATTTTTAATCAATTATTAATTTT 2211

Qy  2341 TTTAGAAATAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2212 TTTAGAAATAATAATATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTT 2271
```

```
Qy  2401 TTTTTCTTTTTTAAAGAGAAAGCAGGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2272 TTTTTCTTTTTTAAAGAGAAAGCAGGCTTTCTCAGACTTTAAAGAATAAAGTCTTTGGGAG 2331

Qy  2461 GTCCCACGGTGTGAGAGAGAGAGCTTGAGGCGACCACCCGCCACCCGACCAAAATTCACCCAGAGAGGGAAAT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2332 GTCCCACGGTGTGAGAGAGAGAGCTTGAGGCGACCACCCGCCACCCGACCAAAATTCACCCAGAGAGGGAAAT 2391

Qy  2521 CTGGTCGGAAGGACCACTCACGGCAGTTCTCGGATCACCTGTATGTCAACGAGAGGGATA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2392 CTGGTCGGAAGGACCACTCACGGCAGTTCTCGGATCACCTGTATGTCAACGAGAGGGATA 2451

Qy  2581 CCGTCCCTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATACACCCATT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452 CCGTCCCTTGAAGAGAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCATACACCCATT 2511

Qy  2641 TCTCTTTGCTTCACAGGTTTTAAACTGTTTTTGCATACTGCTATATAATTCTCTGTCT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512 TCTCTTTGCTTCACAGGTTTTAAACTGTTTTTGCATACTGCTATATAATTCTCTGTCT 2571

Qy  2701 CTCTCTGTTTATCTCTCCCCTCCCCTCCCCTCCTTCTCCCATCTCCATCTCCATTTCTTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2572 CTCTCTGTTTATCTCTCCCCTCCCCTCCCCTCCTTCTCCCATCTCCATCTCCATTTCTTTTGAA 2631

Qy  2761 TTTCCTCATCCTCCAATCCGTATCCGATTCAATCCGTATCCGATTCACAACGCAAAGCA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632 TTTCCTCATCCTCCAATCCGTATCCGATTCAATCCGTATCCGATTCACAACGCAAAGCA 2691

Qy  2821 GTGCTCTGAGTATCACATCACACAAGAGCGAAACACACAACCAGCCTCAA 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692 GTGCTCTGAGTATCACATCACACAAGAGCGAAACACACAACCAGCCTCAA 2751

Qy  2881 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATCAACCAAACAAAGAAAAAATTCCAATG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752 CTTACACTTGGTTACTCAAAAGAACAAGAGTCAATCAACCAAACAAAGAAAAAATTCCACAATG 2811

Qy  2941 AGGAAAACAGGAACCCACCAACACAAGAAACCAATCAACCAAACAAAGAAAAAATTCCACAATG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812 AGGAAAACAGGAACCCACCAACACAAGAAACCAATCAACCAAACAAAGAAAAAATTCCACAATG 2871

Qy  3001 AAAGAATGTATTTTGTCTTTTGCATTTTGTCTTTTTGCATTTGTATAAGCCATCAATATTCAGACAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2872 AAAGAATGTATTTTGTCTTTTGCATTTTGTCTTTTTGCATTTGTATAAGCCATCAATATTCAGACAAAATG 2931

Qy  3061 ATTCCTTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932 ATTCCTTTCTTTAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCC 2991

Qy  3121 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAACACACAGAGATTTTGTATAAGCTGACTCCAG 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2992 CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAGAAGCTACCTCAG 3051

Qy  3181 GTGTTTTTACCTCAGCACCTTGCTCTTGTCTTGTTTTCCTTAGAGATTTTGTAAGCTGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052 GTGTTTTTACCTCAGCACCTTGCTCTTGTCTTGTTTTCCTTAGAGATTTTGTAAGCTGATAG 3111

Qy  3241 TTGGAGCATTTTTTATTTTTTATTTTTAATAAAAATAAGATATCAACT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112 TTGGAGCATTTTTTATTTTTTATTTTTAATAAAAATAAGATATCAACT 3171

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGACAGTGTGCAACAGCTGTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGACAGTGTGCAACAGCTGTTCTCCGCT 3231

Qy  3361 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
         |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  3232 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283
```

```
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-8

Query Match          92.1%;     Score 3143;  DB 4;   Length 3283;
Best Local Similarity 96.2%;     Pred. No. 0;
Matches 3283         Conservative 0;   Mismatches 0;   Indels 129;   Gaps 1;
```

```
Qy     1 GGCACGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCTCCGGGG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGCACGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCGGAGGAGGCTCCGGGG   60

Qy    61 AAGAGACGGATGATGATGAACAAGCTTTACATCGGGAAACAAGCTTTACATCGGGAAACAAGCTTTACATCGGGAACCTGAGCCCGTCACCCCGAC  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGAGACGGATGATGATGAACAAGCTTTACATCGGGAAACAAGCTTTACATCGGGAACCTGAGCCCGTCACCCCGAC  120

Qy   121 GACCTTCCGGCAGCTCTTTGGGGACAGGAAGCGTGCCCCTGCCCCTGCGCGGACAGGTCCTGCTGAAG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACCTTCCGGCAGCTCTTTGGGGACAGGAAGCGTGCCCCTGCCCCTGCGCGGACAGGTCCTGCTGAAG  180

Qy   181 TCCGGGCTACGCCTTCGTGGACTACACCCGGACTACACCCGGGCGCGCCCGGTAC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCCGGGCTACGCCTTCGTGGACTACACCCGGACTACACCCGGGCGCGCCCGGTAC  240

Qy   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAAATCATGGGAAAATCATGGGAAAATGAAATTCAGATTACTCAGTCTCTAAA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGGAAAATGAAATTCAGATTACTCAGTCTCTAAA  300

Qy   301 AAGTTAAGGAGCAGGAAAATTCAGATTCAATATGGGACAGTGGGAATGGGACAGTGAGTCAACAGAC  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGTTAAGGAGCAGGAAAATTCAGATTCAATATGGGACAGTGGGAATGGGACAGTGAGTCAACAGAC  360

Qy   361 TTGGATGGACTTTTGGCTCAATAGCGGGCATCAGTTGAGAACGTCACATATGGGCATCAGTTGAGAACGTCACATATGCCATG  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGATGGACTTTTGGCTCAATAGCGGGCATCAGTTGAGAACGTCACATATGCCATG  420

Qy   421 ACAGAAACCGCGTTGTCAACGTCACATATGCAACAAGAAGAGAAAATAGCCATG  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACAGAAACCGCGTTGTCAACGTCACATATGCAACAAGAAGAGAAAATAGCCATG  480

Qy   481 GAGAAGGTAAGCGGGCATCCCCTTCCCGCCCCCAGCGCGGGGACCACTCTTCCCGG  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGAAGGTGAGCGGGCATCCCCTTCCCGCCGGTGGGGACCACTCTTCCCGG  540

Qy   541 GAAGAGGTGAGCTCCCCTTCCCGCCCCCAGCGCCCTGGGGACTCCCGCTGCGG  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAGAGGTGAGCTCCCCTTCCCGCCCCCAGCGCCCTGGGGACTCCCGCTGCGG  600

Qy   601 GAGCAAGGCCACGCCCTGGGGCCCACTTTCTCAGGCCAGACAGATTCGATTTCCCGCTGCGG  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCAAGGCCACGCCCTGGGGCCCACTTTCTCAGGCCAGACAGATTCGATTTCCCGCTGCGG  660
```

```
Qy   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAG  720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATCCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGGGGCTTGACCATAAAG  720

Qy   721 AACATCACTAAGCACAGACCAGTCCCGGTAGATATCCATAGAAAAGAGACTCTGGAGCT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AACATCACTAAGCACAGACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT  780

Qy   781 GCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCATCCGCATG  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCAGAGAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCATCCGCATG  840

Qy   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGCCGAAAACTAGCCGAAGAGATTCCTCTG  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ATTCTTGAAATCATGCAGAAAGAGGCAGATGAGCCGAAAACTAGCCGAAGAGATTCCTCTG  900

Qy   901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AAAATCTTGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTG  960

Qy   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC  1020

Qy  1021 ATATACAACCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATATACAACCGGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCT  1080

Qy  1081 GAGATAGAGATTATGAAGAGTCGCGTGAAAATGATATGTGGCTGTGTTAAC  1140
           ||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  1081 GAGATAGAGATTATGAAGAGTCGCGTGAAAATGATATGTGGCTGTGTTAAC  1140

Qy  1141 CAACAAGCCAATCTGATCCAGGTTGAACCTCAGCGCCACTTGCATCTTTCAACAGGA  1200
           
Db  1141 ---------------------------------------------------------  1140

Qy  1201 CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCCCCCGCTCGCCCCTACCAC  1260
           
Db  1141 ---------------------------------------------------------  1140

Qy  1261 CCCTTCACTACCACCACTCTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCG  1320
           |||||||||||||||||   |||||||||||||||||||||||||||||||||
Db  1141 -----ACCACTCCGATACTTCTCCAGCCTGTACCCCATCACCAGTTGGCCCG  1191

Qy  1321 TTCCCGCATCATCACTCTTATCCAGCAGGAGATTGTGAATCTTTCATCCCACCCAG  1380
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  1192 TTCCCGCATCATCACTCTTATCCAGCAGGAGATTGTGAATCTTTCATCCCACCAG  1251

Qy  1381 GCTGTGGGCGCCATCATCGGGAGAAGAAGGGCACACATCAAACAGCTCGGCGAGATTCGCC  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252 GCTGTGGGGCGCCATCATCGGGAGAAGAAGGGGCACACATCAAACAGCTCGGCGAGATTCGCC  1311

Qy  1441 GGAGCCTCATCAAGCATTGCCCCTGTTCAAGGCCCAGGACGATGGTCATC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312 GGAGCCTCATCAAGCATTGCCCCTGTTCAAGGCCCAGGACGATGGTCATC  1371

Qy  1501 ATCACCGGGCACCCGGAAGCCCAGTTCAAGGCCCAAAGAGAAGAAAACTGTGAAA  1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372 ATCACCGGGCACCCGGAAGCCCAGTTCAAGGCCCAAAGAGAAGAAAACTGTGAAA  1431

Qy  1561 GAGGAAAACTTCTTTAACCCCGGGTGATTGGCAAGACCGTGAAGCGCATATCAGAGTGCCC  1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432 GAGGAAAACTTCTTTAACCCCGGGTGATTGGCAAGACCGTGAAGCGCATATCAGAGTGCCC  1491

Qy  1621 TCTTCCACAGCTGCCCGGGTGATTGGCCTTCGCGCTGGGGACCACTCTTTCCCGAAC  1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492 TCTTCCACAGCTGCCCGGGTGATTGGCCTTCGCGCTGGGGACCACTCTTTCCCGAAC  1551

Qy  1681 TTAACCAGTGCGAAAGTCATCGTGCCTCCTGACCAAACGCCCAGAGTGAAAATGAGGAAGTG  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552 TTAACCAGTGCGAAAGTCATCGTGCCTCCTGACCAAACGCCCAGAGTGAAAATGAGGAAGTG  1611

Qy  1741 ATCGTCAGAATTATCGGGCACTCTTTGCTAGCCAGACTGCACAGGGATCCAGGGAA  1800
           
```

```
Db  2692  GTGCTCTGAGTATCACATCACACAAAAGGAACACAGAGTCAATGGTACTGTCCTAGCGTTTGGAAG  2751
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2881  CTTACACTTGGTTACTCAAAAGAACAGAGTCAATGGTACTGTCCTAGCGTTTGGAAG  2940

Db  2752  CTTACACTTGGTTACTCAAAAGAACAGAGTCAATGGTACTTGTCCTAGCGTTTGTGGAAG  2811
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2941  AGGAAAACAGGAACCCACCAACCAACCAACCAAAGAAAAATTCCACAATG  3000

Db  2812  AGGAAAACAGGAACCCACCAACCCACCAACCAACAAAGAAAAATTCCACAATG  2871
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3001  AAAGAATGTATTTTGTCTTTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060

Db  2872  AAAGAATGTATTTTGTCTTTTGTCATTTTGGTGTATAAGCCATCAATATTCAGCAAAATG  2931
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3061  ATTCCTTTCTTTAAAAAAAAAAAATGGGAGAGAAGTAGAAATTTACCCAAGGTTGTTGGCC  3120

Db  2932  ATTCCTTTCTTTAAAAAAAAAAAATGTGGAGGAAGTAGAAATTTACCCAAGGTTGTTGGCC  2991
                |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3121  CAGGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAAACACACAGAGAAGCTACCTCAG  3180

Db  2992  CAGGGGCGTTAAATTCACAGATTTTTTTAACGAGAAAAAACACACAGAGAGCTACCTCAG  3051
                ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181  GTGTTTTTACCTCAGCACCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG  3240

Db  3052  GTGTTTTTACCTCAGCACCCTTGCTCTGTGTTTCCCTTAGAGATTTGTAAAGCTGATAG  3111
                ||||||||||||||||||||||||||||||||||||||||||||||
Qy  3241  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3300

Db  3112  TTGGAGCATTTTTTATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAAATAAGATATCAACT  3171
                |||||||||||||||||||||||||||||||||||||||||||
Qy  3301  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCGTGTTCTCGAATTGCTCTCCGCT  3360

Db  3172  GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCGTGTTCTCGAATTGTCTCTCCGCT  3231
                |||||||||||||||||||||||||||||||||||||||||||
Qy  3361  AGCCAAGAACCNATATGGCCCTTCTTTGGACAAAACCTTGAAAATGTTTATTT  3412

Db  3232  AGCCAAGAACCNATATGGCCCTTCTTTTGGACAAAACCTTGAAAATGTTTATTT  3283
```

RESULT 5
US-09-643-597-347
; Sequence 347, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Fan, Liqun
;   APPLICANT: Kalos, Michael D.
;   APPLICANT: Bangur, Chaitanya S.
;   APPLICANT: Hosken, Nancy
;   APPLICANT: Fanger, Gary R.
;   APPLICANT: Li, Samuel X.
;   APPLICANT: Wang, Aijun
;   APPLICANT: Skeiky, Yasir A.W.
;   APPLICANT: Henderson, Robert A.
;   APPLICANT: McNeill, Patricia D.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.455C11
;   CURRENT APPLICATION NUMBER: US/09/643,597
;   CURRENT FILING DATE: 2000-08-21
;   NUMBER OF SEQ ID NOS: 369
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 347
;   LENGTH: 1740
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-643-597-347

Query Match   20.9%; Score 712; DB 4; Length 1740;
Best Local Similarity 64.5%; Pred. No. 5.2e-189;
Matches 1156; Conservative 0; Mismatches 555; Indels 81; Gaps 3;

```
Qy   73  ATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCCCGTCACCGCCGACGACCTCCGCGAG  132
```

```
Db  1612  ATCGTCAGAGATTATCGGGGCACTTCTTTGCTGCTAGCCAGACTGCACGACGCGCAAGATCAGGGAA  1671
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1801  ATTGTACAACAGGTGAAGCAGCAGCAGGAGCAGGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1860

Db  1672  ATTGTACAACAGGTGAAGCAGCAGCAGGAGCAGGAAATACCCTCAGGGAGTCGCCTCACAGCGC  1731
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1861  AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACACGGATGAATGTAGCCCTTCCAACAC  1920

Db  1732  AGCAAGTGAGGCTCCCACAGGCACCAGCAAAACACGGATGAATGTAGCCCTTCCAACAC  1791
                ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1921  CTGACGAATGAGACAAACGCAGCCAAACTCGGACCAGATCGGAGCAAACCAAAGACCATCTGAG  1980

Db  1792  CTGACAGAATGAGACAAACGCAGCCAAACTCGGACCAGATCGGAGCAAACCAAAGACCATCTGAG  1851
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1981  GAATGAGAAGTCTGCGCGGAGCGGCGGCCAGGGGACTCTCGCCGAGAACCCCAGGGGC  2040

Db  1852  GAATGAGAAGTCTGCGCGGAGCGGCGGCCAGGGGACTCTCGCCGAGAACCCCAGGGGC  1911
                |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041  CGAGGAGGGGGGGAAGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC  2100

Db  1912  CGAGGAGGGGGGGAAGTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCC  1971
                ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2101  CCCCAGGGGCTTCTGCGAGGCTTCAGCCATCACTTCACCACTCGGATCTCTCCTGAA  2160

Db  1972  CCCCAGGGGCTTCTGCGAGGCTTCAGCCATCACTTCACCACTCGGATCTCTCCTGAA  2031
                ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2161  CTCCCAGCGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2220

Db  2032  CTCCCAGCGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGC  2091
                |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2221  AAAATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA  2280

Db  2092  AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA  2151
                |||||||||||||||||||||||||||||||||||||||||||||
Qy  2281  GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT  2340

Db  2152  GGGAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGGTAATATAT  2211
                |||||||||||||||||||||||||||||||||||||||||||
Qy  2341  TTTAGAAATAATATCAAATAACTCAACTAACTCCAATTTTTAATCAATTATTAATTTT  2400

Db  2212  TTTAGAAATAATATCAAATAACTCAACTAACTAACTCCAATTTTTAATCAATTATTAATTTT  2271
                |||||||||||||||||||||||||||||||||||||||||||
Qy  2401  TTTTCTTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460

Db  2272  TTTTCTTTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2331
                ||||||||||||||||||||||||||||||||||||||||||||
Qy  2461  GTCTTCACGGTGTGTAGAGAGGACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGGAAAT  2520

Db  2332  GTCTTCACGGTGTGTAGAGAGGACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGGAAAT  2391
                |||||||||||||||||||||||||||||||||||||||||||||
Qy  2521  CTCGTCGGAAGGACACTCACGCGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA  2580

Db  2392  CTCGTCGGAAGGACACTCACGCGCAGTTCTGGATCACCTGTATGTCAACAGAAGGGATA  2451
                ||||||||||||||||||||||||||||||||||||||||||||||
Qy  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGTCATGCCTGTCTAGCTTCATACACCCATT  2640

Db  2452  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTGTCATGCCTGTCTAGCTTCATACACCCATT  2511
                ||||||||||||||||||||||||||||||||||||||||||||||
Qy  2641  TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATCTGCTATATAATTCTCTGTCT  2700

Db  2512  TCTCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATCTGCTATATAATTCTCTGTCT  2571
                ||||||||||||||||||||||||||||||||||||||||||||||
Qy  2701  CTTCTCTGTTTATCTCCCATCCCATCTCAATCTCCATTCTTCTCCATTCTTTTTGAA  2760

Db  2572  CTCTGTTTATCTCTCCCATCCCATCTCAATCTCCATCTCCATTCTTTTTGAA  2631
                ||||||||||||||||||||||||||||||||||||||||||||
Qy  2761  TTTTCCTCATCCCTCCATCTCCATCCCGTATCACGGACCACCCCCCCCCCCAGGCAAAGCA  2820

Db  2632  TTTTCCTCATCCCTCCATCTCCATCTCCGTATCACGGACCACCCCCCCCCCCCCAGGCAAAGCA  2691
                ||||||||||||||||||||||||||||||||||||||||||||
Qy  2821  GTGCTCTGAGTATCACATCACAC  2880
```

RESULT 6
US-09-542-615A-347
; Sequence 347, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C8
; CURRENT APPLICATION NUMBER: US/09/542,615A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 347
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-542-615A-347

Query Match   20.9%;   Score 712; DB 4; Length 1740;
Best Local Similarity   64.5%;   Pred. No. 5.2e-189;

Matches 1156; Conservative 0; Mismatches 555; Indels 81; Gaps 3;

```
Qy    73 ATGAACCAAGCTTACATCGGGAACCTGGGCCCGCCGTCACCGGCGACCTCCGGCAG 132
         ||||||||||||| |||| |||  ||  || || |  || || |  ||||||||||
Db     1 ATGAACCAAACTGTATATGCGGAAACCTCAGCGAGAACGCGGCCCCCTCGGACCTAGAAAGT 60

Qy   133 CTCTTTGGGACAGGAGAGTCGCCCCTGGCGGGGACAGGTCCTGCTGAAGTCGGCTACGCC 192
         |||||   || |   ||||||  | |||||||||||
Db    61 ATCTTCAAGGACGCGCCAAGATCCCGGTGTCGGGACACCTTCTCGGTGGTGAAGACTGGCTACGCG 120

Qy   193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGGCGCCATCGAGACCCTCTCGGGTAAA 252
         |||||||||||  ||||||||||||||||| ||||||||||||
Db   121 TTCGTGGACTGCCCCGGACGAGAGCCGTCAAGGCCATCGGGCCCTTCAAGGCGCGCGCTTTCAGGGTAA 180

Qy   253 GTGGAATTGCATGGGAAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGGAGC 312
         |||| |     |  |  |    |||| |||| || ||||| | | ||
Db   181 ATAGAACTGACGCGGAAACCCATAGAAGTTGAGCACTTCCGGTCCCAAAAAGGCAAAGAATT 240

Qy   313 AGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAGTGGAGGTGTTGGATGACTT 372
         || |||  | || ||  ||||||  |   | |||||
Db   241 CGGAAACTTCAGATACGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 300

Qy   373 TTGGCTCAATATGGAACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAACCGCC 432
         |||  || |   | || || || ||  | || ||||  ||  |   ||||||
Db   301 CTAGTCCAGTATGGAGTGGTGGAGAGCGTGAGCAAGTGAACACTGACCTCGGAAACTGCA 360

Qy   433 GTTGTCAACGTCACATATGCAAGAGAAGAGACAAAAATAGCCATGGAAGAAGCTAAGC 492
         |||||  ||  |  || |  ||  ||| ||| ||  ||||
Db   361 GTTGTAAAATGTAACCTATTCCAGTAAGGACCAAGCTCAGTTCAAGGCTCAGGGAAGAAATTTATGGAAAAATTAAAGAA 420

Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGC 552
         ||||||||||| |||| || | || ||||| | |
Db   421 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATAATCCCTGATGAAACGGCCGCC 480

Qy   553 TCCCCTTGGCCCCCAGCGAGTCGGGGACCAGCAGTGGGGACCACTCTTTCCCGGAGCAAGGC 609
         |||     |  |  |   |
Db   481 CAGCAAAACCCCTTGCAGCAGCCCGTGCCGCCGGGCTTGGGCAGAGGGGCTTCCTCA 540

Qy   610 -----CACGGCCCTGGGGGCACTTCGGAGATTGATTTCCGCTGCGGATC 663
         |     ||  ||| | |||| ||  || | |  ||
Db   541 AGGCAGGGGTCTCCAGAGATCCGTATCCAAGCAGAAACCCATGTGATTTGCCTCTGCGGCTG 600

Qy   664 CTGGTCCCACCCAGTTGTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAAC 723
         ||||||||||||||||||||| |||| ||||||||||||
Db   601 CTGGTTCCCACCCAATTTGTTGGAGCCATCATAGGAAAGAGAGGCTGCCACCATTCGGAAC 660

Qy   724 ATCACTAAGCAGACCAGTCCCGGTAGTATCCATGAAAAGAGAACTCTGGAGTCGCA 783
         |||| | ||||||||   |||  | || |||||
Db   661 ATCACCAAACAGACCCCAGTCTCTAAAATCCACCGTGCACCCAGTGGGGCAAAATCAGAGAGGCTCGCAAAGTCGACCAC 720

Qy   784 GAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATT 843
         |||||  |  |||  |  || ||  ||  ||  ||| ||
Db   721 GAGAAGTCGATTACTACTATCCTCTACTCCTGAAGGCACCTCTGCGGCTTGTAAGTCTATT 780

Qy   844 CTTGAAATCATGCAGAAGAGAGCCAGAGATGAGACCCAAACTAGCCGAAGAGATTCCTCTGAAA 903
         ||||||   |  |   || |  | ||  |||  ||
Db   781 CTGGAGATTATGCATAGGAAGGCTCAGAGCTCAGATATAAAATTCACAGAGGAGATCCCCTTGAAG 840

Qy   904 ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAGGAGGGCAGAAATTTGAAG 963
         |||  ||  || ||   |   || |||| | |
Db   841 ATTTTAGCTCATAATAACTTGTTGGACGTCTTTATTGGTTAAAGAAGAAGAAATCTTAAA 900

Qy   964 AAAATTGAACATGAAACAGGGACCAAGATAACAATCTTCATCTTTGCAGATTTGAGCATA 1023
         |||| || || | ||||  | |   || ||  |   ||||  ||
Db   901 AAAATTGAGCAAGCAGACACAGAACACTAAAATCACCGATATCTCCATTGCGAGGAATTGACCCTG 960

Qy  1024 TACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCTGTGCCAGTGCTGAG 1083
         ||| || ||| | |||||  ||||| |||
Db   961 TATAATCCAGAACGCCACTATTACAGTTAAAGGCAATGTGCCAAAGTGAG 1020

Qy  1084 ATAGAGATTATGAAGAAGCTGCCTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAA 1143
         || |  | |||| || || | |   ||||   | |||| || |||| || |  |  | ||
Db  1021 GAGGAGATCATGAAGAAGAAAATCAGGGAGTCTTATGAAAATGATATTGGCTTTCATGAATCTT 1080
```

```
Qy  1144 CAAGCCAATCGATCCCAGCGGTTGAACCTTCAGCGCACTTGGCATCTTTCAACAGGAGACTG 1203
         ||||  |  |||   || |  | |  |||  |||  | |  |||  |  | |  |
Db  1081 CAAGCACATTTATTCCTGGATTAAAATCTGAACGCCTTGGGTCTGTTCCCACCCACTTCA 1140

Qy  1204 TCCGTGCTATCTCCACCAGCAGCAGGCGTGAATTCGGGAGCTCCCCCGCTGCCCCCTACCACCCCC -------- 1263
         |||   |  ||||||||||  |  |||
Db  1141 GGGATGCCACCTCCCACCTCAGGGGCCCCCTTCAGCCATGACTCCT-------- 1185

Qy  1264 TTCACTACCACCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTC 1323
         |||           |
Db  1186 TT------------------------CCCTACCCGCAGTTTGAGC----- 1204

Qy  1324 CCGGCATCACTCTTATCCAGACAGGAGATTGTGAATCTTCATCCCAACCCAGGCT 1383
         |      | || ||| | | || |
Db  1205 ------------AATCAGAAACGGGAGACGTGTTCATCTGTTTATCCCAGCTCTATCA 1248

Qy  1384 GTGGGCGCCATCATCGGGAAGAAGAGGGGCACACATCAAACAGCTGGCGAGATTCGCCGGA 1443
         |||||||||||||||||||||| ||||| ||| ||||| |||  | |  | |
Db  1249 GTCGGTGCCATCATCGGCAAGCAGGGCCAGCATCATCAAGCAGAGCTTTCTCGGCTTTGCTGGA 1308

Qy  1444 GCCTCTATCAAGATTGCCCCTCGGAAGGCCCAGAGCTCAGCGACGATGGTCATCATC 1503
         |||||||||||||||| ||| |  | ||||| ||| |   |||| | | | |||||
Db  1309 GCTTCAATTAAGATTGCTCAGCGGAAGCCAGAGCACCAGAGTGCTAAAGTGAGGATGATTATC 1368

Qy  1504 ACCGGGCACCGGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGAG 1563
         |||||   | || ||  |||||||||||||  |||||    || | | |
Db  1369 ACTGACCACCAGAGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAGAAAATTAAAGAA 1428

Qy  1564 GAAAACTTCTTTAACCCAAAGAAGAGTGAAGCTGGAAGGCCGCATATCAGATGCCCCTCT 1623
         |||||||| || ||||| ||||| ||||| | ||
Db  1429 GAAAACTTTGTTAGTCCTAAAGAAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCC 1488

Qy  1624 TCCACAGCTGCGCCGGTGATTGGCAAGGTGCAAGAGGTGCAAGACCGTGACGAACTGCAGAACTTA 1683
         ||||| | || | ||| | || | || | || |  |
Db  1489 TTTGCTGTCTGGGAGAGTTATTGGAAAAGGAGCAAAACGGTGAAGTGAATGAACTTCAGAATTTG 1548

Qy  1684 ACCAGTCGAGAAGTCATCGTGCCTCGTCGTGACCAAAACGCCAGATCCGGTAAAATGAGGAAGTGATC 1743
         || |||| |||| | |  ||  || || || ||  ||
Db  1549 TCAAGTGCAGAAGTTGTTGTCCCTGTGTCCCTGTGACCAGACACCTGATGAGGAATGAGACCAAGTGGTT 1608

Qy  1744 GTCAGAATTATCGGGACACTTCTTTGCTTGTAGCCAGATCAGGGAAATT 1803
         |||||||| |||||||||| |  |   | |||| |
Db  1609 GTCAAAATAACTGGTCACTTCTATGTTGCCAGGTGCCAGAGAAAAATTCAGGAAATT 1668

Qy  1804 GTACAACAGGTGAAGCAGCAGAGGAGCAGAATACCCTCAGGGAGTGCCTCAC 1855
         || ||| | |  |||| |||| | ||
Db  1669 CTGACTCAGGTAAAGCAGGTCTCGCAAAGTGGACCAC 1720
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-606-421B-347

Query Match    20.9%;    Score 712;    DB 4;    Length 1740;
Best Local Similarity   64.5%;    Pred. No. 5.2e-189;
Matches 1156;   Conservative   0;   Mismatches  555;   Indels  81;   Gaps   3;

Qy     73  ATGAACAAGCTTTACATCGGGACCTGGACCTCAGAGCCCGTCACCGCCGACGACTTCCGGCAG  132
               ||||||||||||||| | |||| || ||||||||||| |||| ||||||||||||
Db      1  ATGAACAAACTGTATATCGGGAAACCTCAGCGAGAACGCGCCCCCTCCGACCTAGAAAGT   60

Qy    133  CTCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGCTCCGGCTACGCC  192
               ||||| | |||||||| || ||||| || || |||||
Db     61  ATCTTCAAGGACACGGACGCCAAGATCCCGGCTGTCGGGACACCCTTCCTGGTGAAGACTGGCTACGCG  120

Qy    193  TTTCGTGGACTACCCCGACCAGCAACTGGGCCATCCGCGCCATCGAGACCCTCTCGGGTAAA  252
               |||| ||||| || | |||||| |||||||||||| |||||||
Db    121  TTTCGTGGACTGCCCCGGGACCAGAGTGGTGGAGAGCTGTGAGCAAGCTGAAACTGCA  180

Qy    253  GTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAGCTAAGGAGC  312
               || |||||| |||||| |||||||| || |
Db    181  ATAGAACTGCACGGGAAACCCATAGAAGTTGAGCACTTGGTCCCAAAAAGGCAAAGATT  240

Qy    313  AGGAAAATTCAGATTGAAACATCTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTT  372
               |||||| ||||| | | |||| || || | || || |
Db    241  CGGAAACTTCAGATACGATGAAATATCCCGCCTCATTTACAGTCGGAGGAGGTGCTGGAATAGTTTA  300

Qy    373  TTGGCTCAATATGGGACAGTGGACAAGTGGACAAGTCAACACAGACACAGAAAACCGCC  432
               || || || ||||||| || ||| |||||| |||| ||||| | ||| || |
Db    301  CTAGTTCCAGTGATGATGGGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTCGGAAACTGCA  360

Qy    433  GTTGTCAAGTCACATATGGCAACACAGAGAAGCAAAAAATAGCCAGTGGAGAGCTAAGC  492
               |||| || ||||||| |||| || || ||| || | | |||
Db    361  GTTGTAAATGAACTATTCCAGTAAGGACCAAGCTAGACAAGCTAGACAACTGAAT  420

Qy    493  GGGCATCAGTTTGAGAACTACTCCTCAAGATTTCCTACACATCCCGGATGAGAGGTGAGC  552
               ||| ||| ||| || || || |||||||||||||
Db    421  GGATTTCAGTAGTAGAATTTCACCTGAAGTAGCCTATATCCTGATGAAACGGCCGCC  480

Qy    553  TCCCCTTCGCCCCCTCAGCGAGCCCCAGCGTGGGAGACCACTCTTCCCGGGAGCCAAGGC---  609
               || ||| |||||| || |||||| ||| ||||
Db    481  CAGCAAAAACCCCTTGCAGCAGCCCCAGTCTAAAATGCATTGATTTCCCGCTGCGGATC  540

Qy    610  ------CACGGCCCTGGGGGCACTTCTCAGGCCACAGCAGATTGATTTCCCGTGCGCCATC  663
                     ||||||| || |||||| || |
Db    541  AGGCAGGGGTCTCCAGGATCCGTATCCGTATATCAGGAAACCATGTGATTTGCCTCTGCGCCCTG  600

Qy    664  CTGGTCCCCACCCAGTTGTTGGTGCTTGGTTCGGAAGACTGATTCGGACCATAAAGAAC  723
               |||| ||||| |||||| || | || |||
Db    601  CTGGTTTCCCCCAATTTGTTGGACCATCATAGGGACGTGCCACCATTCGGAAC  660

Qy    724  ATCACTAAGGCAGAACATGAAACAGGGACCCAGTCCCGGGTAGAATATCCATAGAAAGAG  783
               || |||| | | ||||||||||| || ||
Db    661  ATCACCAAACAGAACCCAGTCTAAAATGTCCACCGATGATGCATGGGGGCGGCTGCT  720

Qy    784  GAGAGCCTGTCACCATCCATGCCACCAGAGGGGACTTCTGAAGCATGCCGCATGAATT  843
               || ||| || |||||| |||
Db    721  GAGAAGTCGATTACTATCCTCTCACTCCTGAAGGCACCTCTGAAGCTGTAAGTCTATT  780

Qy    844  CTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAA  903
               || |||||| ||| | |||||| ||| |
Db    781  CTGGAGATTATGCATAAGGAGCTCAAGATATAAAAATCACAGAGAGAATTGGAAG  840

Qy    904  ATCTTGGCCACAATGCTTGGTTCGGAAGACTGATTGGAAGAGAGGCAGAAATTTGAAG  963
               || |||||| || | ||||||| || | |||| || |||||
Db    841  ATTTTAGTCATAATAACTTGTTGGACGTCTTATTGGTTAAAGAGGAAGAAATCTTAAA  900

Qy    964  AAAATTGAACATGAAACAGGGACCAAGATAACAATCCATCTTTGCAGGATTTGAGCATA  1023
               |||||||||| || ||| || | | ||||
Db    901  AAAATTGAGCAGAAGCAACAACAACTAAAATCACGATATCTCCATTGCAGGAATTGACGCGCTG  960

Qy   1024  TACAACCCGGAAAGAACCATCACTGTAGAAGGCACAGTTGAGGCCTGTCCAGTGCTGAG  1083
```

```
;  NUMBER OF SEQ ID NOS: 8
;  SEQ ID NO 4
;  LENGTH: 4159
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-061-709-4

Query Match      20.8%;    Score 710.4;  DB 3;     Length 4159;
Best Local Similarity   64.5%;    Pred. No. 2.7e-188;
Matches 1155;  Conservative   0;  Mismatches 556;   Indels   81;   Gaps    3;
```

```
Qy      73  ATGAACAAGCTTTACATCGGGAACCTGGGACGCCCGTCACCGCCGAGACCTTCCGGCAG  132
            |||||||||  ||||||||||||| |||  |||||||||  ||||| ||||| ||| |||
Db     251  ATGAACAAACTGTATATCGGGAACCTCAGCGAGACGAAACGCCGCCCTCGGACCTAGAAAGT  310

Qy     133  CTCTTTGGGACAGAGAAGCTGCCCCTGGCGGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
            |  ||| |||||||||  |||||||| |||  |||||||||  |||||||||||||| |||
Db     311  ATCTTCAAGACGACGGCCAAGATCCCCGTGTCGGGACCCTTCCTGGTGAACGACTGGCTACGCG  370

Qy     193  TTCGTGGACTACCCCGACCAGAACTGGGCCATCGCGCCATCGAGACCCTCTCGGGTAAA  252
            |||||| ||||||||||||| |||| ||||||||||  ||||||||  |||||||||||
Db     371  TTCGTGGACTGTCGCCCCGAACGAGAGTCGGGGCCTCAAGGCCATCGAGGCCGCTTTCAGGTAAA  430

Qy     253  GTGGAATTGCATGGAGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGTAAGGAGC  312
            ||||||| ||||||  ||| || || |||||||||  ||||||| |||||||| |||||
Db     431  ATAGAACTCACGGGAAACCTAGAAGTTGAGCACTTCCGTCCCAAAAAGGCAAAGGAATT  490

Qy     313  AGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTCGGAGGTGTTGGATGGACTT  372
            |   ||||||| |||||| ||| ||| ||||||| |||  ||||||| | |||||  |||
Db     491  CGGAAAACTTCAGATACGAAATATCCCGCCTCATTACAGTAGTGGAGGTGCTGGATAGTTTA  550

Qy     373  TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAGAAACCGCC  432
            |||||||||||||| ||||||||  |||||||||||  || |||||||| || ||||
Db     551  CTAGTCCAGTATGGAGTGTGTGAGCAAGTGAACACTGACTCCGGAAACTGCA  610

Qy     433  GTTGTCAACGTCACATATGTCACAACAGAGAAGAAAAATAGCCATGGAGAAGCTAAGC  492
            ||||| ||||| |||| || || |||| ||||||||||| || ||||| || |||||||
Db     611  GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGCTGACAAGCACTAGACAAACTGAAT  670

Qy     493  GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGC  552
            |||||||||||| | ||| ||||| ||||||||||| || ||||| || ||||||||||
Db     671  GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy     553  TCCCCTTCGCCCCCTGGCGAGCCCAGCGTGGGGACCACTCTCTCCCGGGAGCCAAGGC  609
             ||||| || || ||||||||||||||||  ||||||||| || || ||||| ||||
Db     731  CAGCAAAACCCCTTGCAGCAGCCCCGAGGTCCGCCGGGGCCTTCGCCCGGGGCCTCCTCA  790

Qy     610  ------CACGCCCCTGGGGGCACTTCTCAGGCCCAGACAGATTGATTTCCCGTTGCCGATC  663
                  |||||||  ||||| |||||| |||||||| || || || ||  |||||||||||
Db     791  AGGCAGGGGTCTCCAGGATCCGTATCCAAGGACCAAGCAGCGAAACCATGCTGATTTGCCTCTGCCCG  850

Qy     664  CTGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAAC  723
            ||||| ||||| |||| || ||| ||||| || ||||| |||| |||| ||| |||||
Db     851  CTGGTTCCCCAATTTGTTGGAGCCCATCATAGGAAAAGAAGAGTGCCACCATTCGGAAC  910

Qy     724  ATCACTAAGCAGACCAGTCCCGGGTAGTATATCCATAGAAAAGAAGACTCTGGAGCTGCA  783
            |||| |||| |||||  |||  ||| ||  ||| |||| ||||| ||||| || |||||||
Db     911  ATCACCAAACAGACCCAGTCTAAAATGATGTCCACCGTAAAGAAAAGAAGATGCGGGGCT  970

Qy     784  GAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGATGATT  843
            |||||||||||||||| ||| || ||| ||||| |||| ||  |||||||| |||||  ||
Db     971  GAGAAGTGATTAATACTCTCTTCTACTCCTGAAGGCCACCTTCTGCGCGTTTGTAAGTCTATT  1030

Qy     844  CTTGAAATCATGCAGAAGAGAGGCAGATGAGACCAAAACTAGCCGAAGATTCCTCTGAAA  903
            || |||||| ||||||||||  ||| |||||| ||| ||||| || ||||||| |||||
Db    1031  CTGGAGATTATGCATAAGGAAGCTCAAGATATAAAATTCACAGAGAGATCCCCTTGAAG  1090

Qy     904  ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAGCAGAAATTTGAAG  963
            ||||| || |||  || ||| ||||| |||| ||  |||||||| |||||||||||
Db    1091  ATTTTAGCTCATAATAACTTCGTTGGACGTCTTATTGGTAAGAGAGGAAGAAATCTTAAA  1150

Qy     964  AAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA  1023
```

```
Db    1151  AAAATTGAGCAAGACAAGACAGACACTAGACTAAAATCACGATATCTCCATTGCAGGAATTGACGGCTG  1210
            ||||||||| || ||||| |||||||||| |||||| ||  ||| || |||||||  |||||||||||

Qy    1024  TACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGGCCAGTGCTGAG  1083
            ||  ||||| | |||||  ||||||||  || ||||  |||||||| ||| ||||| ||
Db    1211  TATAATCCAGAAGCGACTATTACAGTTAAAGGCAATGTTGAGACATGTGCAAAGTGAG  1270

Qy    1084  ATAGAGATTATGAAGAAGCTGCGTGAGGCCTTGAAAAATGATATGTGGCTGTTAACCAA  1143
            |||||||||||||||||||||| ||  |||||||| ||  |||||||||||| | |||
Db    1271  GAGGAGGATCATGAAGAAGAAATCAGGAGTCTTATGAAAATGTATATTGCTTCTATGAATCTT  1330

Qy    1144  CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGACTTGGCATCTTTCAACAGGACTG  1203
            |||| ||| |||| |  |||  |||||  |||  || |  || | || ||| ||| |||
Db    1331  CAAGCACATTTAATTCTGGATTAAATCTGAACGCGCCTTGGGTCTGTCTCCCACCCACTTCA  1390

Qy    1204  TCCGTGTATCTCCACCAGCAGAGGGCCCGGGAGTCCCCGCGTGCCCCTACCACCCC  1263
            || ||| ||||||  |||| ||||| ||||||  ||| ||
Db    1391  GGGATGCCACCTCCCACCTCAGGGCCCCTTCAGCCATGACTCCT------------  1435

Qy    1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTC  1323
                                           |||||||||||| ||||||
Db    1436  ------------------------------CCCTACCCGCAGTTTGAGC------  1454

Qy    1324  CCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCAACCCAGGCT  1383
                                          |||||||| ||| ||  |||||| |||||
Db    1455  ------------------------AATCAGAAACGGAGACTGTTCATCAGTTTATCCCAGTCTTATCA  1498

Qy    1384  GTGGGGCCATCATCGGAAGAAGAGGGCCACACATCAAACAGCTCGGCGAGATTCCCCGGA  1443
            ||| |||||||||  |||| |||||||||||||||| ||||  || ||||| ||| |||
Db    1499  GTCGGTGCCATCATCCGGCAAGCAGGGCCAGCAGCACCAAGAGCTTTCTCGCTTTGCTGGA  1558

Qy    1444  GCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGCGTCAGCGAAGGATGGTCATCATC  1503
                 ||| ||||||| ||||| || ||||||||||  |||||||| ||||||||| |||
Db    1559  GCTTCAATTAAGATTGCTCACGCCGAAGCAGGACCAGTGCTAAAGTGAGGATGATGATTATC  1618

Qy    1504  ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGACGATCTTCAAGGCGAAGAG  1563
            || ||| |||||  || ||||||| ||| ||||||||||| ||| ||| |||||||
Db    1619  ACTGGACCACCAGGACCTCAGTTCAAGGCTCAGGAAGAGAATTTATGGAAAAATTAAAGAA  1678

Qy    1564  GAAAACTTCTTTAACCCAAAGAAGAGTGAAGCTGGAAGGCGCATATCAGAGTGCCCTCT  1623
            |||||| |||||| || |||| |||| |||| ||||| ||| |||| ||||||| |||
Db    1679  GAAAACTTTGTTAGTCTAAAGAAGAAGTGAAACTTGAACTCATATCAGATGCCATCC  1738

Qy    1624  TCCACAGCTGCCGCGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTGCAGAACTTA  1683
            || |||||| || ||  |||| ||  ||||| || ||  ||||| |||  || ||||||
Db    1739  TTTGCTGCTGCAGAGTTATTGGAAGAGGAGCAAAATCGGTGAATGAACTTCAGAATTTG  1798

Qy    1684  ACCAGTGCAGAAGTGCATCGTGCGTCCTGCCTCGTCGGTGACCACCGAATGCAAACGCCCAGATC  1743
            || ||||| |||||| |||| ||  || ||||||||||| |||| |||| ||| |||||||
Db    1799  TCAAGTGCAGAAGTTGTGTCCCGTGCACCTGATAGCAGACTGCAGACTGCACCAAGTGGTT  1858

Qy    1744  GTCAGATTATCGGGACTTCTTTGCTAGCCAAGAGAAC  1803
            || ||| ||| || ||| ||| || |||| ||||||||
Db    1859  GTCAAAATAACTGGTCACTTCTATGCTTCTGCCAGGTTGCCCAGAGAAAAAATTCAGGAAATT  1918

Qy    1804  GTACAACAGTGAAGCAGTGAAGCAGCAGAATACCCTCAGGGAGTCGCCTCAC  1855
            || ||| || |||| |||| || ||||| |||| ||||||||| |  ||||||
Db    1919  CTGACTCAGGTAAAGCATGTAAAGCTCTGCAAAGTCTGCGACCAC  1970
```

```
RESULT 9
US-09-899-651-4
;  Sequence 4, Application US/09899651
;  Patent No. 6576756
;  GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
```

; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 4
;   LENGTH: 4159
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-899-651-4

Query Match   20.8%;   Score 710.4; DB 4;   Length 4159;
Best Local Similarity   64.5%;   Pred. No. 2.7e-188;
Matches 1155; Conservative   0;   Mismatches 556;   Indels 81;   Gaps   3;

RESULT 10
US-09-643-597-175
; Sequence 175, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;   APPLICANT: Wang, Tongtong

Query Match   20.8%;   Score 710.4;   DB 4;   Length 4181;

Best Local Similarity   64.5%;   Pred. No. 2.8e-188;
Matches 1155;   Conservative   0;   Mismatches 556;   Indels 81;   Gaps 3;

```
Qy     73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGGTCACCGCCACGACCTCCCGCAG  132
          ||||||| |  || ||||  |||||||| |||||  | ||  ||| | ||||| | | |||
Db    251 ATGAACAAACTGTATATCGGAAACCTCAGCGAGAACCGCGCCCCCTCGGACCTAGAAAGT  310

Qy    133 CTCTTTGGGACAGGAGAGTGCCCTGCCGGCGGACAGGTCCTGCTGAAGTCCGGCTACGCC  192
          || |||||||| |||   || || ||| | |||| |||  ||| |||| |||| ||| |||
Db    311 ATCTTCAAGGACGCGCAAGATCCCGGTGTGCGGGACCCTTCCTGGTGAAGACTGGCTACGCG  370

Qy    193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCCGCGCGCCATCGAGACCCCTCGGGTAAA  252
          |||||||||| ||| | |  ||||  ||||  | | |  | ||| | | | |||  || |||
Db    371 TTCGTGGACTGCCCGGACGAGAGCTGGGCCCTCAAGGCATCGAGGCGCTTTCAGTGAA  430

Qy    253 GTGGAATTGCATGGGAGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC  312
          ||| ||  | |||  |  ||| |||| ||| |  ||  ||||  ||||  || | ||||||
Db    431 ATAGAACTGCACGGAGAAACCCATAGAAGTTGAGCACTCGGTCCCAAAAAAGGCAAAGGATT  490

Qy    313 AGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTTGGATGGATGACTT  372
          || ||||| |||||  |||| ||||| |||||| |||||||| || | |||| | || |||
Db    491 CGGAAACTTCAGATACGAAATATCCCGGCCTCATTTACAGTCGGGAGTGCTGGATAGTTTA  550

Qy    373 TTGGCTCAATATGGGACAGTGGAGATGTGGAAGATGTGAACAGTCAACACAGACAGAAACCGCC  432
          ||| |||  | |||||||  || |  ||| | |  ||||  || |  ||| ||  ||| ||||
Db    551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGACACTGACTCGGAAACTGCA  610

Qy    433 GTTGTCAAGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGAGAAGCTAAGC  492
          |||||||||| ||| | || | ||| |||||| ||  | | |  ||| | | |||||
Db    611 GTTGTAAATGTAACCTATTCCAGTAGGACCAAGCAGCACTAGACAAACTGAAT  670

Qy    493 GGGCATCAGTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGGTGAGC  552
          ||  || |||| | |||  ||  |||| |||||  | | ||||| || | ||||  |||
Db    671 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy    553 TCCCCCTTGCCCCCTCAGCGAGCCCCAGGCGTGGGACCCACTCTTCCCGGGAGCAAGGC--  609
          || ||| | || | ||||| | ||   ||| |  ||  | || |||||| ||| ||||
Db    731 CAGCAAAACCCCTTGCCAGCCCCGAGGTCCGGAATCCCTGATGAAATGGCCGCC  790

Qy    610 ------CACGGGCCCTGGGGGCACTTCTCAGCAGATTGATTTCCCGCTGCGGATC  663
                |||| | || ||| | |||  |||| | | | ||  |  ||| | | |||
Db    791 AGGCAGGGGGTCTCCAGGATCCGTATCCAAGCAGAAACCATGTGATTTGCCTCTGCGCCTG  850

Qy    664 CTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGACGGCTTGACCATAAAGAAC  723
          ||||| || |||||||||| |||  | |  ||  |||| | || |||| ||| | |||||
Db    851 CTGGTTCCCACCCAATTGTTGGAGCATCATAGGAAAAGAGGGTCCACCATTCGAAC  910

Qy    724 ATCACTAAGCAGAACCCAGTCCCGGGTAGATATCCATAGAAGAAAGAGAACTCTGGAGCTGCA  783
          ||  ||||||| ||||||| | || |||  | ||  | |  ||| ||||  ||| ||||||
Db    911 ATCACCAAACAGACCCAGTCTAAAATCGTATGTCACCGTAAAGAAATGCGGGGGCTGCT  970

Qy    784 GAGAAGCCTGTCACCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATGATT  843
          ||||| |||  || ||| | |  |||| | || || |||||||  | | || |||
Db    971 GAGAAGTGCGATTACTATCCTCTCTCCTGAAGGCACCTGTGAAGGTTGTAAGTCTATT  1030

Qy    844 CTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAGCAGCCAGAGATTCCTCTGAAA  903
          || || ||| | |||| |  | | |  | |||  ||| |  || | | | ||||||
Db   1031 CTGGAGATTATGCATAAGGAAGCTCAAGATAACATTCACGAGACTATAAAAATTTGAAAG  1090

Qy    904 ATCTTGCCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGGAGGCAGAAATTTGAAG  963
          || ||| |  | | || |||| |   |||  ||  |  ||||| || ||| | ||| |||
Db   1091 ATTTTAGCTCATAATAACTTTGTTGGACGTCTTATTGGTGGTAAAGAAGGAAGAAATCTTAAA  1150

Qy    964 AAAATTGAACATGAAACATGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA  1023
          |||| |||| || || | ||| |  ||| || | | | || |||||||| ||| || |||
Db   1151 AAAATTGAGCAAGAACAGACACTAAATCACGATATCTCCATTCAGGAATTGACGCGCTG  1210

Qy   1024 TACAACCCGGAAAGACCATCACTGTGAAGGGCACAGTTGAGGCCTGTCGCCAGTGCTGAG  1083
          || || | |||||| || | ||| ||||| || || |||||||  | ||||| ||| |||
Db   1211 TATAATCCAGAACGACGACTATTACAGTTAAGGCAATGTTGAGACATGTGCCAAAGCTGAG  1270

Qy   1084 ATAGAGATTATGAAGAAGCTGCGTGAGCCGTTTGAAAATGATATCGTCTGGCTGTTAACCAA  1143
          ||||| || ||  | |  |  |   || |   |||||| ||||  | |  ||||  |||||
```

```
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(4181)
;  OTHER INFORMATION: n=A,T,C or G
US-09-480-884A-175

Query Match          20.8%;   Score 710.4;  DB 4;   Length 4181;
Best Local Similarity 64.5%;   Pred. No. 2.8e-188;
Matches 1155;  Conservative  0;  Mismatches 556;  Indels 81;  Gaps  3;

Qy     73 ATGAACAAGCTTACATGGGAACCTGAGCCCCGTCACCGGCGACGACCTCCGGCAG 132
          |||||||||||||||||||||  ||||  |||| |  | ||| | ||||||  ||
Db    251 ATGAACAAACTGTATATCCGAAACCTCAGCGAGACGCGCCCCCTCCGACCTAGAAAGT 310

Qy    133 CTCTTGGGACAGGAGAGCTGCCCTGGGGGACAGCAGTGCTCGTGAAGTCCGGCTACGCC 192
          ||||||  ||| || || |||| ||| |  |  | | |  ||  |  ||| | || |
Db    311 ATCTTCAAGGACGCCAAGATCCCGTGCGGGGACCCCTCCTGGTGACGACTGCTACGACCG 370

Qy    193 TTCGTGGACTACCCCGACCAGAACTGGGCCATCGAGACCCTCTCGGGTAAA 252
          ||||||||||||  |  ||| |||||| | |||| || |||||||| ||||
Db    371 TTCGTGGACTGCCCCGACGAGCAGCTTCCTCAAGGCCATCAGGTCTTTCAGGTAAA 430

Qy    253 GTGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTCTCTAAAAAGTAAGGAGC 312
          ||||| | ||  || | |||| |   ||| | |  | | | || |||| || |||
Db    431 ATAGAACTGCACGGGAAACCCATAGAGTTGAGCACTGTCCGTCCCAAAAAGGCAAAGGATT 490

Qy    313 AGGAAAATTCAGATTGAAACATCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTT 372
          |||||||  | ||| |||| | |  |  | | |||  || |||| | | | |  ||
Db    491 CGGAAACTTCAGATACGGAAATATCCCGCCTCATTTACAGTGGGATGGTGCTGGATAGTTTA 550

Qy    373 TTGGCTCAATATGGGACAGTGGGAGAATGTGGAGAAGAGCAACAAGTGGAACAAACTGGAAGTGGGAGGAGAGCTAAGC 432
          |||||  ||  | || |||| |||  | | |||| || | |||| | | ||  |  |||| ||||| | |
Db    551 CTAGTCCAGTGAGTGGTGGAGAGCTGTGGAGAAGTGAACAAGTGAACACTGACTGGGGAAACTGCA 610

Qy    433 GTTGTCAACGTCACATATGCAACAAGAAGAAGAAGCAACAAAAAAATAGCCATGGAAGAGAAGAGCTAAGC 492
          | | |||||  | ||| |  | || ||||| || || || | | || ||||| | || | |  ||||
Db    611 GTGTAAATGTAACCTATCCAGTATGACCTAAGCACTAGACAAGCACTAGACAACTGAAT 670

Qy    493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGC 552
          ||| | |  || ||||| | | ||| ||| | |  | || |  ||||  |||||| |
Db    671 GGATTCAGTTAGAGAATTCACCTTGAAAGTAGCCTATATCCTGATGAATGGCCGGC 730

Qy    553 TCCCCTTCGCCCCCTCAGCGAGCCAGCGTGCGGGGACCACTCTTCCCGGGAGCAAGGC--- 609
          ||||| ||| | | || ||| |  ||| |  ||| |  ||  | || ||| ||||
Db    731 CAGCAAAACCCCTTGCACGCACCCCGAGTCGCGCGGGGGCTTGGGCAGGAGGGGCTCCTCA 790

Qy    610 ------CACGGCCCTGGGGCACTTCTCAGGACCAGCAGATGATTTCCCGCTGCGGATC 663
                |  | | | ||| | || ||| |||  ||| |||||| ||||| ||| |||||
Db    791 AGGCAGGGGTCTCCAGGATCCGTATCCAAGCAGCAGAAACCATGTGATTTTGCTCTGTCGCCCTG 850

Qy    664 CTGGTCCCCACCCAGTTGTTGTTGGTGCCATCATCGGAAAGGAGGGGCTTGACCATAAAGAAC 723
          ||||| |||  | ||  | |  | || || ||   ||| || | | ||  || || ||| |
Db    851 CTGGTTCCCACCCAATTTGTTGGAGCCATCATCATAGGAGGAAGAGAGGGTGCCACCATTCGGAAC 910

Qy    724 ATCACTAAGCAGAACCCAGTCCCGGTAGTATATCCATAGAAAGAAGAACTCTGAGCTGCA 783
          ||| | || ||| |||| ||| | ||||| ||||  | ||||| || | |   || ||
Db    911 ATCACCAAACAGACCCCAGTCTAAAATCGATGTCAAAATGATGTCTAAAGAAAATGCGGGGGCTGCT 970

Qy    784 GAGAAGGCCTGTCACCATCCATGCAACCCCAGAGGGGACTTCTGAAGCATGCGCCATGATT 843
          |||||| | | | || |   || ||| |  | || |  | | | ||  | | ||||| |
Db    971 GAGAGTGCATTCGATTACTATATCCTCTACTCCTGAAGGCAACCCTTGCGGCTTGTAAGTCTATT 1030

Qy    844 CTTGAAATCATGCAGAAAGAGGCAGATGAGAACCAAACTAGCCGAAGAGATTCCTCTGAAA 903
          | ||| || | | |   || || | | || |  || || || || | || | ||||| | ||
Db   1031 CTGGAGATTATGCATAGGGAAGGAAGTCAAGGATATAAAATTCACAGAGAGATCCCCTTGAAG 1090

Qy    904 ATCTTGGCACACAATGGCTTGTTGGAAGACTTGGTTGGAAGACTGAAATTTGAAG 963
          || | | | | | || |||| |||| || | | || | |||  ||| |||| |||
Db   1091 ATTTTAGCTCATAATAACTTTGTTGGACGTCTTATTGTGAAGAAGAAGAAATCTTAAA 1150

Qy    964 AAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGAGCATA 1023
          ||||||||||| |||| ||||| ||||| |||| || |  ||||||| |
Db   1151 AAAATTGAGCAAGAACAGACAAGACACTAAAATCACGATATCTCCATTGCAGGATTGACGCTG 1210
```

(continued on reverse / rotated sequence block)

```
Db   1271 GAGGAGATCATGAAGAAATCAGGGAGTCTTATGAAAAATGATATTGCTTCTATGAATCTT 1330

Qy   1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTG 1203

Db   1331 CAAGCACATTTAATTCCTGGATTAAATCTGAACGGCTTGGGTCTGTTCCCACCCACTTCA 1390

Qy   1204 TCCGTGCTATCTCCACCAGCAGGCAGGGCCCCGGGAGTCCCCCGTGCCCCCTACCACCCC 1263

Db   1391 GGGATGCCACCTCCCACCTCCAGGGCCCCCTTCAGCCATGACTCCT------------ 1435

Qy   1264 TTCACTACCCCTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTGCGCCGTTC 1323

Db   1436 --------------------------CCCTACCCGCAGTTTGAGC--------- 1454

Qy   1324 CCGATCATCACTCCTTATCCAGAGACCAGGAGGATTGTGAATCTCTTCATCCCAACCCAGCT 1383

Db   1455 -----------AATCAGAAACGGAGACTTGTTCATCAGTTATCCCAGCTCTATCA 1498

Qy   1384 GTGGGGCCGTCATCATCGGGAAGGAGGGGGCACACATCAAACAGTCGGCAGGATTCGCCGGA 1443

Db   1499 GTCGGTGCCATCATCGGGCAAGCAGGGGCCAGCACATCAAGCCAGCTTTCTCGGCTTTGCTGGA 1558

Qy   1444 GCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGCAGTCGCCGAAAGGATGGTCATCATC 1503

Db   1559 GCTTCAATTAAGATTGCTCCAGCCGGAAGCCAAGCACCACAGATGGCTAAAGTGGAGGATGGTGATTATC 1618

Qy   1504 ACCGGGCCACCGGAGCCCCAGTTCAAGGGCACGGACGGATCTTCTTGGGGAAACTGAAAGAG 1563

Db   1619 ACTGGACCACCAGAGCCTCAGCTTCAAGGGCTCAGGTTCAAGGGAAGAGAAATTTATGGAAAAAAATAAAGAA 1678

Qy   1564 GAAAACTTCTTTAACCCCCAAAGAGAAGAAGTGAAGCTGGAGACGAATCATCAGAGTGCCCTTCT 1623

Db   1679 GAAAACTTTGTTAGTCCTAAAGATGAACTTGAAGCTCATATCAGAGTGCCCATCC 1738

Qy   1624 TCCACAGCTCGGCCGGGTGATTGGCAAGGTGGCAAGAACCGTGACCGCTGACAACTTA 1683

Db   1739 TTTGCTGCTGCAGAGTTATTAGTGGCAGAGTGATTGGAAAAGGAGCCAAAACGTGAATGAGTTCAGAATTTG 1798

Qy   1684 ACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAGGCCCAGATGAAAATGAGGAAGTGATC 1743

Db   1799 TCAAGTGCAGAAGTTGTGTCCCTCGTGACCAGACACTTGGTGAGAATGACCCAAGTGGTT 1858

Qy   1744 GTCAGAATTATGGGACACTTCTTTGCTAGCCGACAGCGCCAAGATCAGGGAAATT 1803

Db   1859 GTCAAAAATAACTGGTCACTTCTATGCTTCTAGCTGCCCAGAGAAAAATTCAGGAAATT 1918

Qy   1804 GTACAACAGTGAAGCAGCAGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCAC 1855

Db   1919 CTGACTCAGTGAAAGCAGCAGCACCAACAACAGAGAGGGCTCTGCAAGTGGACCAC 1970
```

```
Qy  1024  TACAACCCGAAAGAACCATCACTGTGAAGGCACAGTTGAGGCCTGTGCCAGTGCTGAG  1083
          ||||   |||||   |||||||||||||| |||||||  |||||||||||||||||||
Db  1211  TATAATCCAGAACGACCACTATTACAGTTAAAGGCAATGTTGAGACATGTGCCAAAGCTGAG  1270

Qy  1084  ATAGAGATTATGAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAA  1143
          ||||   ||||   |||||||||||||||||||||||||||||||||||||||   ||||
Db  1271  GAGGAGATCATGAGAAAATCAGGGAGTCTTATGAAAATGATATTGCTTCTATGAATCTT  1330

Qy  1144  CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTCAACAGGACTG  1203
          ||||   ||||   |||||||||||||||||||||||||||| ||| |||  |||||||
Db  1331  CAAGCACATTTAATTCCTGGATTAAATCTGAACGCCTTGGGTCTGTTCCCACCCACTTCA  1390

Qy  1204  TCCGTGCTATCTCCACCAGCAGGGGCCCCGGGAGCTCCCCCGCTGCCCCTACCACCCC  1263
          |||  |||    |||||||||||||||||  ||||| |||
Db  1391  GGGATGCCACCTCCCACCTCCCAGGGGCCCCCTTCAGCCATGACTCCT----------  1435

Qy  1264  TTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTTC  1323
                                         -----CCCTACCCGCAGTTTGAGC------  1454
Db  1436  ---------

Qy  1324  CCGCATCATCACTCTTATCCAGAGCGATGATTGTGAATCTCTTCATCCCACCCAGCGCT  1383
                                      |||||||||||||||||||
Db  1455  -----------AATCAGAAAACGGAGACTGTTCATCAGTTTATCCCAGTCTCTATCA  1498

Qy  1384  GTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACACAGCTGGCCGAGATTCGCCCGGA  1443
          ||||  ||    |||||||||||||||||||  |||||||  |||| |||| |||  |||||
Db  1499  GTCGTGCCATCATCGGCGACAGCAGGCCAGCAGCTTTCGCGTTGCTGGA  1558

Qy  1444  GCCCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGAGGATGGTCATCATC  1503
          |||    |||||||||||||     ||||||||||||||   |||||||||| |||||||
Db  1559  GCTTCAATTAAGATTGCTCCAGCGGAAGCACCAGCATGCTAAAGTGAAGATGGTGATTATC  1618

Qy  1504  ACCGGGCCACCGGAAGCCCAGTTCAAGGCCCAGGGACCGATCTTTGGGAAACTGAAAGAG  1563
          |||||||||||||||||||   ||||||||||||| |||||| |||  ||||||| ||||
Db  1619  ACTGACCACCAGAGCGGCTCAGTTCAAGGCTCAAGGTCAGGGAAGAATTTATGAAAAATTAAAGAA  1678

Qy  1564  GAAAACTTCTTTAACCCAAAGAAGAGAAGTGAAGCTGGAAGCCGCATATCAGAGTGCCCTCT  1623
          |||||||  ||   || |||||||| |||   |||  ||||| ||  ||||||||||||||
Db  1679  GAAAACTTTGTTAGTCCTAAAGAAGAGTGAAACTTGAAGCTCTCATATCAGAGTGCCATCC  1738

Qy  1624  TCCACAGCTGGCCGGGTGATTGCCAAAGGTGCCAAGACCGCAGACGCTGAACGAACTGCAGAACTA  1683
          |||||| ||||||||||||   |||    ||| |||| |||||||| ||||| |||||||||||
Db  1739  TTTGCTGCTGGCAGAGTTATTGAAAAAGGAGGTGAAAAACGGTGAATGAACTTCAGAATTTG  1798

Qy  1684  ACCAGTGCAGAGTCATCGTGCCTCGTTGCTCCGTGTCCCTGACCAAACGCCCAGATGAAAATGAGGAAGTGATC  1743
          |||||| |||||||||||||| ||  ||   ||||||   ||| ||||| |||   |||||||||||| ||||
Db  1799  TCAAGTGCAGAAGTTGTGTCCCCGTGTGACCCACCACCTGATGAGGAATGACCAAGTTGTT  1858

Qy  1744  GTCAGAATTATCGGGCACTTCTTTGCTACCAGCAGGCCAAGATCAGGGAAATT  1803
          |||   |||   ||||||||||||||||||||||||||||| |||||||||||
Db  1859  GTCAAAATAACTGGTCACTTCTATGCTTCTATGAATCTTCAAGGCGCCATCAGGAAATT  1918

Qy  1804  GTACACAGTGAAGCAAGCAGCAGGAGCAGAGAAATACCCTCAGGGAGTCGCCTCAC  1855
          ||||||||||| | |  |||| ||||  |||||    ||||
Db  1919  CTGACTCAGTAAAGCAGCAGCACCAACAACAGAAGGCTCTGTGGGCCCTTCAAAGTGCCGACCAC  1970
```

```
Query Match             20.8%;  Score 710.4;  DB 4;  Length 4181;
Best Local Similarity   64.5%;  Pred. No. 2.8e-188;
Matches 1155;  Conservative    0;  Mismatches  556;  Indels  81;  Gaps  3;

Qy   73  ATGAACAAGCTTTACATCGGGAACTTGAGCCCCGCGTCACGCCCACGACCTCCCGCGCAG  132
         |||||||||| ||||| |||||||||| |||||||||||   ||||||||||   |||||
Db  251  ATGAACAAACTGTATCGGAAACCTCAGCGAGACAGCGCCGCCCCTCGGACCTAGAAAGT  310

Qy  133  CTCTTTGGGACAGGAAGCTGCCCCTGGCGGGACAGTGTCCTGCTGAAGTCCTGCGGCTACGCC  192
         |||||| ||||||||| ||||  || ||||||||||||| |||| |||||||| |||||||||
Db  311  ATCTTCAAGGAGCGCCAAGATCCCGGTGTCGGGAGACCCTTCCTGGTGAAGACTGGCTACGCG  370

Qy  193  TTCGTGGACTACCCCGACCAGACTGACGAACTGGGCCATCGGCCTCGGGTAAA  252
         ||||||||| ||||||||||||| |||  ||||||||  ||   |||||||||
Db  371  TTCGTGGACTGCCCGACGAGAGTCGGGCCCTCAAGGCCATCAAGGCGCTTTCAGTAA  430
```

```
Db   1455 ----------AATCAGAAACGAGGACTGTTCATCAGTTTATCCCAGTCTCTATCA 1498
Qy   1384 GTGGGGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCGACTTGCCCGGA 1443
Db   1499 GTCGGTGCCATCATCGGCAAGCAGGGCCACGGGCGGAAGCACCAGATGCTAAAGTGAGGATTTCTCGCTTTGCTGGA 1558
Qy   1444 GCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGCGTCGCGAAAGGATGGTCATCATC 1503
Db   1559 GCTTCAATTAAGATTGCTCCAGCGAAGGCACCAGATGCTAAAGTGAGGATGGTGATTATC 1618
Qy   1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGGCCAGGACGGATCTTTGGAAACTGAAAGAG 1563
Db   1619 ACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAGAATTTATGGAAAAATTAAAGAA 1678
Qy   1564 GAAAACTTCTTAACCCCAAAGAGAGAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCCTCT 1623
Db   1679 GAAAACTTTGTTAGTCCTAAAGAGAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCC 1738
Qy   1624 TCCAGCTGCCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGACGAACTTA 1683
Db   1739 TTTGTGCTGGCTGGCAGAGTTATTGGAAAAGGAGCCAAAACGTGAATGAACTTCAGAATTTG 1798
Qy   1684 ACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAAACGCCAGATGAAAATGAGGAGTGATC 1743
Db   1799 TCAAGTGCAGAGTTGTGTTGTCCCTCGTGACCAGACACCTGACCAGTGCAGTTGGTT 1858
Qy   1744 GTCAGAATTATCGGCGACTCTTGTTGCTAGCCAGACTCGACAGCCCAAGATCAGGGAAATT 1803
Db   1859 GTCAAAATAACTGGTCACTTCTATGCTTGCCCAGTGCCAGAGAAAAATTCAGGAAATT 1918
Qy   1804 GTACAACAGGTGAAGCAGCAGCAGGAGCAGCAGAAATACCCTCAGGGAGTGCGCCTCAC 1855
Db   1919 CTGACTCAGGTAAAGCAGCAGCAACACAGAGGCTCTGCAAAGTGGACCAC 1970
```

RESULT 13
US-09-606-421B-175
; Sequence 175, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; CURRENT FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 4181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3347)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3502)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3506)
; OTHER INFORMATION: n=A,T,C or G

```
Qy   253 GTGGAATTGCATGGGAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGC 312
Db   431 ATAGAACTGCACGGGAAAACCCATAGAAGTTGAGCACTCCGGTCCCAAAAAAAGGCAAAGGATT 490
Qy   313 AGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGAGGTGTTGGATGACTT 372
Db   491 CGGAAAACTTCAGATACGAGAAATATCCGCCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 550
Qy   373 TTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACAAACCGCC 432
Db   551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTCGGAAACTGCA 610
Qy   433 GTTGTCAAGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGGGAGAAGCTAAGC 492
Db   611 GTTGTAAAATGTAACCTATTCCAGTCAGTAAGGACCAACAAGCACTAGACAAACTGAAT 670
Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGAAGAGGTGAGC 552
Db   671 GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC 730
Qy   553 TCCCCTTCGCCCCCCAGCGAGCCCAGCGTGTGGGACCACTCTTCCCGGGAGCAAGGC--- 609
Db   731 CAGCAAAACCCTTGCAGCAGCCCGAGGTCGCCGGGGGTTGGGCAGAGGGGCCTCCTCA 790
Qy   610 -----CACGCCCTGGGGCACTTCTCAGCCCACAGATTGATTTCCCGCTCGGGATC 663
Db   791 AGGCAGGGGGTCTCCAGGATCCGTATCCAAGCAGAGAAAACCATGTGATTTGCCTCTCGCGCCTG 850
Qy   664 CTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGAGAGGGCTTGACCATAAAGAAC 723
Db   851 CTGGTTCCCACCCAATTTGTTGGAGGCATCATAGGAAAAGAAGGTGCCACCATTGGAAAC 910
Qy   724 ATCACTAAGACAGACCCAGTCCCGGGTAGTATCCATAGAAAAGAGAACTCTGGAGCTGCA 783
Db   911 ATCACCACCAAACAGACCCCAGTCTAAAATCGATGTCACCGTAAAGAAAAATCGGGGGCTGCT 970
Qy   784 GAGAAGGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGATT 843
Db   971 GAGAAGTCGATTACTATCCTTCTCTACTCCTGAAGGCCACCTGTGTAAGTCTCTATT 1030
Qy   844 CTTGAAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAA 903
Db   1031 CTGGAGATTATGCATAAGGAAGCTCAAGATATAATAAAATTCACAGAAGAGATCCCCTTGAAG 1090
Qy   904 ATCTTGGCACACAATGCTTGGTTGGAAGACTGATTGGAAAGAAGAGGCAGAAATTTGAAG 963
Db   1091 ATTTTAGCTCATAATAACTTTGTTGGACGTCTTATTGGTAAAGAAGGAGAAATCTTTAA 1150
Qy   964 AAAAATTGAACATGAAACATGGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGACATA 1023
Db   1151 AAAATTGAGCAAGACAGCACAGACAACTAAAATCACGATATCTCCATTGCAGGAATTGACGCGTG 1210
Qy   1024 TACAACCCGGAAAGAAGCAACATCACTGAACGGGCACAGTTGAGGCCTGTTGCCAGTGCTGAG 1083
Db   1211 TATAATCCCAGAACGCACTATTACAGTGAAGGCAATGTTGAGACAGTGTGCCAAAGCTGAG 1270
Qy   1084 ATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGTCGCTGTTAACCAA 1143
Db   1271 GAGGAGATCATGAAGAAAATCAGGAGTCTTATGTGAAAAATGATATTGCTTCTTATGGAATCTT 1330
Qy   1144 CAAGCCAATCGATCCCAGGGTTGAACCTTCAGCGCCACTTGGCATCTCTTTCAACAGGACTG 1203
Db   1331 CAAGCACATTTAATTCCTGGATTAAATCTGAACGCTTGGGTCTTGTTCCCACCCACTTCA 1390
Qy   1204 TCCGTGCTATCTCCACCAGGAGGCCCGGCGAGCTCCCCGGCCATGGGAGCTCCTACCACCCC 1263
Db   1391 GGGATGCCACCTCCCACCTCCAGGGCCCCCTTCAGCCATGACTCCT-------------- 1435
Qy   1264 TTCACTACCCACTCCGGATACTTTCTCCAGCCTGTACCCCCATCACCAGTTTGGGCCGGTTC 1323
Db   1436 ----------------CCCTACCCCCGCCAGTTTGAGC------ 1454
Qy   1324 CCGCATCATCACTCTTATCCAGAGGACAGGAGATTGTGAATCTCTTGAATCCCAACCCAGGCT 1383
```

```
FEATURE:
NAME/KEY: unsure
LOCATION: (3520)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3538)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3549)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3646)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3940)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3968)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (3974)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4036)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4056)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4062)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4080)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4088)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (4115)
OTHER INFORMATION: n=A,T,C or G
US-09-606-421B-175

Query Match             20.8%;
Best Local Similarity   64.5%;
Matches 1155;   Conservative   0;
```

```
Qy  313  AGGAAAATTCAGATTCGAAACATCCCTCTCACCTGCAGTGGGAGGTGTTGGATGGACTT  372
              |||||||||||||||||||||||||||  |||||||    ||||  || ||| ||
Db  491  CCGAAACTTCAGATACGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGCTGGATAGTTTA  550

Qy  373  TTGGCTCAATATGGACAGTGGAGAATGTGAACAAGTCAACACAGACACAGAGAAACCGCC  432
              |||||||||||||||||||||||||||||||    ||||||||||||
Db  551  CTAGTCCAGTAGGAGTGGTGTGAGCAGCTGTGAGCAAGCAAGCACTGACTCCGAAACTGCA  610

Qy  433  GTTGTCAAGTCACATATGCAACAAGAAGAGAAGACAAAAATAGCATGGAGAAGCTAAGC  492
              |||| |||||||||||||||  |||||||||||  |||
Db  611  GTTGTAAATGTAACCTATTCCAGTAAGGACCAAGCTAGACAAGCACTAGACAAAACTGAAT  670

Qy  493  GGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGATGAAGAGTGAGC  552
              |||  |||||||| ||||||||| |   |||||||||||||||||  ||||||||||
Db  671  GGATTTCAGTTAGAGAATTTCACCTTGAAAGTAGCCTATATCCCTGATGAAATGGCCGCC  730

Qy  553  TCCCCTTCGCCCCTCAGCGACCCCAGCCACTCTTCCCGGGAGCAAGGC---  609
              |||||||||||||||||  ||||   ||||||  |
Db  731  CAGCAAAACCCTTGCAGCAGCCCGAGGTCGCCGGGGTCGCCGGAGAGGGCTTCCTCA  790

Qy  610  ------CACGCCCCTGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGCCGATC  663
                    ||||||||||||||  |||||||  |||||||||| ||| |||||||||||||
Db  791  AGGCAGGGGTCTTCCACCCAGTTGTTGGTGCCATCATCGAAAAGAGGGCTTGACCATAAAGAAC  850

Qy  664  CTGGTCCCCACCAGTTGTTGGTGCCATCATCGAAAAGAGGGCTTGACCATAAAGAAC  723
              |||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db  851  CTGGTTCCACCCAATTTGTTGGAGCCATCATAGGAAAAGAGAGGTGCCACCATTCCGAAC  910

Qy  724  ATCACTAAGACACCAGTCCCGGGTAGATATCCATAGAAAGAGAACTCTCGGAGCTGCA  783
              |||||||||||||||||||||||  |||||||||  ||||||||||||||  ||||||
Db  911  ATCACCAAACAGACCCAGTCTAAAATCGATGGTCCACCGTAAAGAAATGCGGGGGCTGCT  970

Qy  784  GAGAAGCCTGTCACCATCCATCGCACCCCAGAGGGGACTTCTGAAGCATGCCCGCATGATT  843
              ||||||||||||||||||||||  |||||||||||||||||||||||||||||||  ||||
Db  971  GAGAAGTCGATTACTATCCTCCTGAAGCCACTCTGCGGCTTGTAAGTCTATT  1030

Qy  844  CTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAATACTAGCCGAAGATTCCTCTGAAA  903
              |||||||||| || ||| ||||||||||||| ||||||||||  |||||||||||||||
Db  1031 CTGGAGATTATGCATAAGGAAGTCAAGATATAAAATTCACAGAAGAGATCCCCTTGAAG  1090

Qy  904  ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAAGCCAGAAATTTGAAG  963
              |||||||||||||||||||||||||||||||  ||  ||||||  |||||
Db  1091 ATTTTAGTCATAATTGTTGGACGTCTTATTGGTAAAGAAGAAGAAATCTTAAA  1150

Qy  964  AAAATTGAACATGAAACCAAGATAACAATCCATCTTGCAGGATTGAGCCATA  1023
              ||||||||||||||||||||||  |||||||||||  |||||||||||
Db  1151 AAAATTGAGCAAGACAAGACACATAAAATCACGATATCTCCATTGCGGAATTGACGCTG  1210

Qy  1024 TACAACCCGGAAGAACCATCACTGTGGAAGGGCACAGTGGCCTGTGCCAGTGCCGCGAG  1083
              |||  ||||||||||||||||||||||  |||||||||||||||||||||||
Db  1211 TATAATCCAGAACGCACTATTACAGTTAAAGGCAATGTTGACACATGTGCCAAAGTCGAG  1270

Qy  1084 ATAGAGATTATGAGAAGCTGCCTGAGCCTTTGAAAAATGATATGTCTGGCTGTTAACCAA  1143
              ||||  ||  |||||||||||||||||||||  ||||||||| |||||||||||  ||||
Db  1271 GAGGAGGATCATGAAGAAATCAGGGAGTCTTATGAAAATGATATTGCTTTCTATGAATCTT  1330

Qy  1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTGGCATCTTTCAACAGGACTG  1203
              |||||||||||||  ||||||||||| || ||||  ||| |||||||||||||
Db  1331 CAAGCACATTTAATTCCTGATTAAATCTGAACGCCTTGGGTCGTTCCCACCCACTTCA  1390

Qy  1204 TCCGTGCTATCTCCACCAGCAGGGCCCCGGAGCTCCCGGCCATGACTCCT----------  1263
              ||| | ||| |  |||||||||||||||||||  |||||| ||||||||
Db  1391 GGGATGCCACCTCCAGGCCCTTCAGGCTGGCTGGTGAGACTGGCTACGCG  1435

Qy  1264 TTCACTACCCACTCCGGATACTTCTCCAGCGTGACCCCATCACCAGTTTGGCCGTTC  1323
              ||||||||||||||||   ||||  ||||||||||||||||||||||||||  |||||
Db  1436 ----------------------------CCCTACCCGCAGTTTGAGC------  1454

Qy  1324 CCGCATCATCACTCTTATCCAGCAGAGATTGTGAATCTTCATCCAACCCAGGCT  1383
              ||||||||||||||||||||||  |||||||||||||||||||  ||||||||||||
Db  1455 ----------AATCAGAAACGAGACTGTTCATCAGTTTATCCAGCTGTTATCA  1498
```

```
Length 4181;
Pred. No. 2.8e-188;   DB 4;
Mismatches 556;   Indels   81;   Gaps   3;

Qy   73  ATGAACAAGCTTTACATGGGACAGCCGAACCTGAGCCCGTCACCGCCGACCTCCGGCAG  132
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  251  ATGAACAAACTGTATATCGGAAAACCTCAGCGAGAACGCGCCCCCTCGGACCTTAGAAAGT  310

Qy  133  CTCTTTGGGACAGGAAGCTGCCCCGGCGGGGACAGTGCCCCTCGCTGCTGAAGTCGCTACGCC  192
              ||||  |||||||||||||| ||| ||||||||||||||| |||  ||
Db  311  ATCTTCAAGGACGCGCCAAGATCCCGGTGTCGGGACCCTTCCTGGTGAAGACTGGCTACGCG  370

Qy  193  TTCGTGGACTACCCCGGACGAGGAAAAATCATGAAGTGATTACTACTCCAGTTCTCTAAAAAGCTAAGGAGC  252
              ||||||||||||||||||||||||||   ||||||||||||||||||| ||||||  ||||||||
Db  371  TTCGTGGACTGCCCGGACGCGAGCTGGCGCCCTCAAGGCCATCGAGCCGCTCAAAAAGAGGCAAGGTAAA  430

Qy  253  GTGGAATTGCATGGGAAAAATCATGAAGTGAGCACTGGATCTTTATCCCAGCTGTTATCA  312
              |||||||||||||||||||||||||||||||||||||||||||  ||||| |  ||||
Db  431  ATAGAACTGCACGGGGAAACCCAGAAGTTGAGCACTGGAGACTGGCACTCGTCCGTCCCAAAAAAGGCAAAGGATT  490
```

```
Qy   371 CGGAAAATACAGATCCGCAATATTCCACCTCAGCTTCCGAGATGGGAGTGCTAGATAGCCTG 430
         |||||||     ||||||||  | |||||||||| |||||||||  ||||||||| |||||||
Db   371 CGGAAAATACAGATCCGCAATATTCCACCTCAGCTTCCGAGATGGGAGTGCTAGATAGCCTG 430

Qy   431 TTGGCTCAATATGGACAGTGGAGAATGTGAGAACAAGTCAACAGAGAACCGCC 432
Db   373 TTGGCTCAATATGGACAGTGGAGAATGTGAGAACAAGTCAACAGAGAACCGCC 432

Qy   433 CTGGCTCAGTGACGTGACAGTGGAGAACTGTGAGCAAGTGAACACTGAAAGTGAGACAGCG 490
Db   431 CTGGCTCAGTGACGTGACAGTGGAGAACTGTGAGCAAGTGAACACTGAAAGTGAGACAGCG 490

Qy   491 GTTGTCAACGTCACATATGCAACAAGAGAGAAGCAAAATAGCCATGGAGAAGTAAGC 492
Db   493 GTTGTCAACGTCACATATGCAACAAGAGAGAAGCAAAATAGCCATGGAGAAGTAAGC 492

Qy   551 GTGGTCAACGTCACCTACTACTTAACGGGAGCAGAACCAGGCAAGCTATCATGATGAAGCTAAAT 550
Db   551 GTGGTCAACGTCACCTACTACTTAACGGGAGCAGAACCAGGCAAGCTATCATGATGAAGCTAAAT 550

Qy   493 GGGCATCAGTTTGAGAACTACTCCTTCAAGATTCCTCAAGATTCCTCATCCGGATGAAGAGGTGAGC 552
Db   553 GGCCATCAACTGGAGAACCATGCCCTGAAGCTCTCCTACATACCTGATGAGCAGGATAACA 610

Qy   553 TCCCCTTC--------GCCCCTCAGCCGGAGCCCAGCGGTGGGACCACTCTTCCCGGGAG 603
Db   611 CAAGGTCCTGGAGAATGGGCGTCGTGGAGGCTTTGGGTCTCGGGGGCCAGCCCCGGCAAGGG 670

Qy   604 CAAGGCCACGCCCCTGGGGCCACTTCTCAGCCGCACAGCAGATTGATTTCCCGCTGCGGATC 663
Db   671 TCGCCCGTGGCGAGCAGGGGCTCCAGCCAGCAGCCAGTGGCGACATCCCTCCTCCGCGGCTC 730

Qy   664 CTGGTCCCCACCCAGTTTGTTGGTGCCATCATCCGGAAAGAGAGGGCTTGACCATAAAGAAC 723
Db   731 CTGGTGCCTACGCAGTGATGTGAGGGCTATCATTGGCAAGGAGGGTGCCACCATCCGAAAC 790

Qy   724 ATCACTAAGCAGACCAGTCCCGGTAGTATCATAGAAAAGAACTCTGGAGCTGCA 783
Db   791 ATCACAAAACAGACCAGTCCAGTCCAAAATAGACGTGCATAGGAAGGAGGAATGCGGGCCTGCG 850

Qy   784 GAGAAGCCTGTCACCATGCCACCCAGAGGGGACTTCTGAAGCATGCCGATGATT 843
Db   851 GAGAAGGCCATCAGCGTGCATTCAACCCCTGAAGCTGCTCCTCCTCCGCGTGCAAGATGATC 910

Qy   844 CTTGAAATCATGCAGAAAGAGGCAGATGACAACTTACGAGCAGAGATTCCTCTGAAA 903
Db   911 TTGGAGATTATGCACAAGAGGCAAGGAACACCGCAAAACCGGCAGATGAAGTTCCCCTGAAG 970

Qy   904 ATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAGAAGCAGAAATTGAAG 963
Db   971 ATCCGGCTCATAACAACTTCGTCGGCGGACTCATTGGCAAGGAGGCCGGAACCTGAAG 1030

Qy   964 AAAATTGACATGAAACAGGAGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATA 1023
Db  1031 AAGGTGGAGCAGGAGCACACAGAGAGATCGAGAGATCACCACCTGAGGCCTGAGGACCTCACGCTC 1090

Qy  1024 TACAACCCGAAAGACCATCACTGTGAGGGACACAGTGAGGCCTGGCCGCCAGTGCTGAG 1083
Db  1091 TATAACCCTGAGAGGCATCACTGTGAGGGGCCCATTGAGAACTGTTGCAGGGCCGAG 1150

Qy  1084 ATAGAGATTATGAAGAAGTGCGTGAGAGCTGCGTGAGCTGTGTGAAAATGATATGCTGGCTGTAACCAA 1143
Db  1151 CAGGAGGATCATGAAGAAAGTCGAGAGGTGCTTACGAGGCGTTGGCCGGCCACTGAGCGTTG 1210

Qy  1144 CAAGCCAATCTGATCCCAGGGTTGAACCTCAGCGCACTTGGCATCTTTTCAACAGGACTG 1203
Db  1211 CAGTCCCACCTCACCTCCTGGGCTTAACCTGGCGTGTGAGGTCTTCTTCCCGCTCCACACCCCC 1270

Qy  1204 TCCGTGCTATCTCCACAGAGGGCCCGGGACAGGTCCTGCGTGAAGTCCGGCTACGCC 1263
Db  1271 AGCGGTGTCCCTCCTCC------TCCCAGCAGTGTCACTGGGGCTGCTCCCTATAGCTCC 1324

Qy  1264 TTCACTACCCACTCCGGATACTTCTGCCGTCCTCGCCCATCACCAGTTTGGCCCGTTC 1323
Db  1325 TTCA----------------------------------------------- 1328

Qy  1324 CCGGATCATCACTCTTATCCAGAGCAGAGATTGTGAATCTTCATCCAACCCAGAGC 1383
Db  1329 --------TGCAGGCTCCGGACGGAGGAGATGGTACAAGTGTTCATCCCCGACCCCAGGCT 1378

Qy  1384 GTGGGCGCCATCATCGGAAGAAGGGGGCACACATCAAACAGCTGGCGAGATTCGCCGA 1443
Db  1384 ...
```

```
Qy  1384 GTGGGGCCCACCATCGGAAGAAGGGGGCACACATCAAAACAGCTGGCCGAGATTCCCGGA 1443
         ||||   |||||||||||| ||||||  |||||||||||
Db  1499 GTCGGTGCCATCATCGGCCAGCAGGCCACGCAGGGCCAGCCATCAGCTCTTCTCGCTTTGCTGGA 1558

Qy  1444 GCCTCTATCAAGATTGCCGTGGGGCTGCCCTGGAAGGCCCAGAGCTCAGCGAAGGATGGTCATCATC 1503
Db  1559 GCTTCAATTAAGATTGCTGCTGCTGCCAGCAGGGCTGTCCAGGAGCCCAGATGGCTAAAGTGCTGAATTATC 1618

Qy  1504 ACCGGGCCACCGGAAGCCCAGTTCAAGGCCGAAGGCGACGATCTCTTTGGGAAACTGAAAGAG 1563
Db  1619 ACTGGAGCACCAGAGGCTCAGTTCAAGGCTCAGGAAGAATTTATGGAAAAATTAAAGAA 1678

Qy  1564 GAAAACTTCTTTAAACCCCAAAGAAGAGTGAAGCTGGAAGCCGCAATATCAGAGTGCCCTCT 1623
Db  1679 GAAAACTTTGTTAGTCCTAAAAAGAAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCCATCC 1738

Qy  1624 TCCACAGCTGCCGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTA 1683
Db  1739 TTTGCTTGCTGGCAGAGTTATTGGAAAAGGAGGCAAAACGGTGAATGAACTTCAGAATTTG 1798

Qy  1684 ACCAGTGCAGAAGTCATCGTCGCCTCGTGACCAAAACCCCAGATGAAAATGAAAATGAGGAAGTGATC 1743
Db  1799 TCAAGTGCAGAAGTTGTTGTCCCTCGTGACCAGACACCTGATGAGAATGACCAAGTGGTT 1858

Qy  1744 GTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATT 1803
Db  1859 GTCAAAATAACTGGTCACTTCTATGCTTGCCAGGTTGCCCAGAGAAGAAAAATTCAGGAAATT 1918

Qy  1804 GTACAACAGGTGAAGCAGCAGGAGCAGAAATACCTCAGGGAGTGCGCCTCAC 1855
Db  1919 CTGACTCAGGTAAAGCAACAGCAGCACCAACAACAGAAGGGCTCTGCAAAGTGGACCAC 1970
```

RESULT 14

US-09-261-855-1

; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus

US-09-261-855-1

Query Match   20.3%;   Score 691.8;   DB 3;   Length 2224;
Best Local Similarity   63.5%;   Pred. No. 2.9e-183;
Matches 1152;   Conservative   0;   Mismatches 582;   Indels 81;   Gaps 3;

```
Qy    73 ATGAACAAGCTTTACATCGGAAAATCATGGGAAAATCATGAAGTTGATTACTCAGTCTCTTAAAAAGCTAAGGAGC 132
Db   131 ATGAACAAGCTTTACATCGGCAACCTTCAACGAGAGTGCAACCAGAGTGACCCCCGCAGACTTGGAGAAA 190

Qy   133 CTCTTTGGGGACAGGAAGCTGCCCTGGCGGGACAGGTCTGCTGGAAGTCCGGCTACGCC 192
Db   191 GTATTCGCGGAGCACAAGAGATCCTACAGCGCCACGCGCCAGTTCTTGGTCAATCGGCTCACGCC 250

Qy   193 TTCGTGGTGGACTACCCACCCGACCAGAACTGGGCCATCTGGCCGATGAAGCCAATCAGGTGAAA 252
Db   251 TTCGTGGATTGCCCCCACGCAGCGCACTCGGCCAGGGCCATCATCGGAAACTTTCTCGGGAAA 310

Qy   253 GTGGAATTGCATGGGAAAATCATGGAAGTGATTACTCAGTTCATGAACAACTCAGTCCCCAAAAAAACAAAGGAGT 312
Db   311 GTAGAACTGCAAGAAAACGTCTAGAGATTGAACAAACTCAGTGTTCATCCCCGACCCCAGGCT 370

Qy   313 AGGAAAATTCAGATTCGAAACATCCCTCCTCACCTGCAAGATGCTGGCAGTGGATGGACTT 372
```

Db 1379 GTGGGGCCATCATTGGCAAGAGGGCCAGCACATCAAACAACTTCTCCGTTTCGCCAGC 1438
Qy 1444 GCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGAGGCCCAGCGGAAAGGATGTGTCATCATC 1503
Db 1439 GCCTCCATCAAGATTGCTCCACCAGAAACACCTGACTCCAAAGTTCGAATGGTCGTCATC 1498
Qy 1504 ACCGGGCCACCGAAGCCCAGTTCAAGGCCCAGGGCGGATCTTTGGGAAACTGAAAGAG 1563
Db 1499 ACTGGACCCCAGAGGCTCAGTTCAGGTTCAAGGCTCAGGGAAGAATTTATGGCAAACTAAAAGAA 1558
Qy 1564 GAAAACTTCTTAACCCAAGAAGAGTGAAGCTGGAAGCGCATATCAGATGTCCCTCT 1623
Db 1559 GAGAATTTCTTTGGTCCCAAGGAGGAAGTAAAGCTAGAGACCACCATACCGGTTCCGGCT 1618
Qy 1624 TCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTTA 1683
Db 1619 TCAGCAGCCGCGCCGCGTCATCGGCAAAGGCGCAAAACCGTGAATGAGCTGCAGAACTTG 1678
Qy 1684 ACCAGTGCAGAGTCATCGTGCCTCGTCGTGACCAAAACGCCAGATGAAAATGAGGAAGTGATC 1743
Db 1679 ACTGCGAGCTGAGGTGTGTAGTGCCAAGAGACCAGACCCCGATGAGAACGAACCAAGTCATT 1738
Qy 1744 GTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCGCCAAGATCAGGGAAATT 1803
Db 1739 GTTAAGATCATCGGACATTTCTTATGCCAGCCAGATGCCTCAGCGGAAGATCCGGACACATC 1798
Qy 1804 GTACAACAGGTGAAGCAGCAGGAGAGAGAAATACCCTCAGGGAGTCGCCTTCACAGCCGCAGC 1863
Db 1799 CTGGCTCAAGTTAAGCAACAGCACCGAGAGGGACAGAGCAACCTGCCCCAGGCACCGGAGG 1858
Qy 1864 AAGTGAGGCTCCCAC 1878
Db 1859 AAGTGACCCCGCCCC 1873

Db 224 TGAGACGGCAGTGGTGATGTCACCTATTCCACCGGGAGCAGCAGCAGCCATCAT 283
Qy 483 GAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACACTACTCCGGATGA 542
Db 284 GAAGCTGAATGGCCACCAGTTGGAGAACCATGCCTGGAAGGTCTCCTACATCCCCATGA 343
Qy 543 ------AGAGGTGAGTGAGCCTC 593
Db 344 GCAGATAGCACAGGGACCTGAGAATGGGCGCGCCGAGGGGCCTTTGGCTCTCGGGGTCAGCC 403
Qy 594 TTCCCGGGAGCAAGGCCACGCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCC 653
Db 404 CCGCAGGGCTCACCTGTGGCAGCGGCGGGCCCCAGCAGCAGCAAGTGACACATCCC 463
Qy 654 GCTGCGGATCCTCCTGGTCCCACCCAGTTGTTGGTGCCATCATCGGAAGAGGAGGCTTGAC 713
Db 464 CCTTCGGCTCCTCGTGCCCACCCAGTATGTGGTGCCATTATTGGCAAGGAGGGGGCCAC 523
Qy 714 CATAAAGACATCACTAAGCACAGCACCAGTCCCGGTAGATATCCATAGAAAAGAGAACTC 773
Db 524 CATCCCGCACACATCACAAAACAGACCCAGTCCAGATAGACGTGCATAGGAAGGAGAACGC 583
Qy 774 TGGAGCTGCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGGACTTCTGAAGCATG 833
Db 584 AGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCTGAGGGCTGCTCTCCGCTTG 643
Qy 834 CCGGCATGATTCTTGAAAATCATCGCAGAAGAGCAGATAGACCCAAACTAGCCGAAGAGAT 893
Db 644 TAAGATGATCTTGGAGATTATGCATAAAAGAGCATAATAAAGAGCTAAGACACCAAAAACGGGCTGACGAGGT 703
Qy 894 TCCTTCTGAAAAATCTTGGCACACAAATGGCTTGGTTGGAAGGACTGATTGGAAAAGAAGGCAG 953
Db 704 TCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACG 763
Qy 954 AAATTTGAAGAAAATTGAACAATGAAACATGAAACAACAGGACCAAGATAACAATCTCATCTTTGCAGGA 1013
Db 764 GAACCTGAAGAGAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGA 823
Qy 1014 TTTGAGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTGC 1073
Db 824 CCTTACCCTTTACAACCTGAGGAGCCATCATCGTGAAGGGGCCATCGAGAATTGTTG 883
Qy 1074 CAGTGCTGAGATAGAGATTATGAAGAGAGCTGCGTGGCGCCTTTGAAAATGATATGCTGGC 1133
Db 884 CAGGGCCCGAGCAGGAGAAATAATGAAGAAAGTTCGGGAGGCCCTATGAGAATGATGCTGGGCTGC 943
Qy 1134 TGTTAACCAACAGCCAATCTGATCCCCAGGGGTGAACCTGAACCTCAGCCGCACTTCAGCGCCACTCTTTTC 1193
Db 944 CATGAGCTCTCAC------CTGATCCCTGGCCTGAACCTGGCTGCTGTAGGGTCTTTTCCC 997
Qy 1194 AACAGGACTGTCCGTGCTATCTCCACCAGAGGGCCCCGGAGCTCCCCCGTCCCCGTCCCCC 1253
Db 998 AGCTTCATCGGCGCAGTCCCGCGCCT------CCCAGCGTTACTGGGGCTGCTCC 1051
Qy 1254 CTACCACCCCTTCACTACCCGGCCATTCTCCAGCGTACCCCGTACCCCCATCACCAGTT 1313
Db 1052 CTATAGCTCCTTTATGC--------------- 1068
Qy 1314 TGGCCCGTTCCCGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTCTTCATCCC 1373
Db 1069 ----------------------AGGCTTCCCGAGCAGGAGATGGTGCAGGGTGTTTATCCC 1105
Qy 1374 AACCCAGGCTGTGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACAGCTGGCCGAG 1433
Db 1106 CGCCCAGGCAGGTGGCCGGGGCCATCATCGGCAAGAAGGGGCAGCAATCAAACAGCTCTCCCG 1165
Qy 1434 ATTCGCCGGAGCCCTCATCAAGATTGCCCTCGGCGGAAGGCTCAGCGAAACGGAT 1493
Db 1166 GTTTGCCAGCGCGCCTCCATCAAGATTGCCACCACCGGAAACACCTGACTCCAAAGTTCGTAT 1225
Qy 1494 GGTCATCATCACCGGGCCCGAAGCCCAGTTCAAGGCCGCCAGGGACGGATCTTTGGGAA 1553

Db 1878 AAGTGAGGCTCCCAC 1878
Db 1873 AAGTGACCCCGCCCC 1873

RESULT 15
US-09-061-709-7
; Sequence 7, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7

Query Match   16.7%;   Score 570.6;   DB 3;   Length 1946;
Best Local Similarity   63.0%;   Pred. No. 2.5e-149;
Matches 997;   Conservative 0;   Mismatches 499;   Indels 87;   Gaps 4;

Qy 303 GCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTCCACCTGCAGTGGGAGGTGTT 362
Db 104 GGTTAGGAGCCGAAAATTCAAATCCGAAATATTCCACCCCAGTCCAGATGGAAGTACT 163
Qy 363 GGATGACTTTGGCTCAATATGGACAGTGAGAATGTGAACAAGTCAACACAGAACAC 422
Db 164 GGACAGCCTGCTGCTCAGTAGTGGTACAGTAGGAACTGTGAGCAAGTGAACACCGAGAG 223
Qy 423 AGAAACCGCCGTGTCAACGTCACATATGCAACAAGAGAAGAAGCAAAAATAGCCATGA 482

```
Db   1226 GGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAA 1285

Qy   1554 ACTGAAAGAGGAAAACTTCTTTAACCCCAAAGAGAAGAAGTGAAGCTGGAAGCGCATATCAG 1613
          ||||  ||||||  |||||  |||||   ||||| |   |||||||||| | | |  | ||
Db   1286 ACTCAAGGAGGAGAACTTCTTTGGTCCCAGGAGAAGTGAAGCTGGGACACCCACATACG 1345

Qy   1614 AGTGCCCTCTTCCACAGCTGGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACT 1673
          ||||   ||| | ||||||| ||||||  |||||  |  |||||  || ||||||||
Db   1346 TGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTT 1405

Qy   1674 GCAGAACTTAACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGA 1733
          ||||||||| ||    ||||| ||| ||  ||| ||  ||||||| |||| |||| || |
Db   1406 GCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGA 1465

Qy   1734 GGAAGTGATCGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGCTGCACAGCGCCAAGAT 1793
          ||||||| | ||| ||||  | | |||||  |||||  ||||| | |||| ||  |||||
Db   1466 CCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGAT 1525

Qy   1794 CAGGGAAATTGTACAACAGTGAAGCAGCAGCAGAAATACCCTCAGGGAGTCGCCTC 1853
          ||||| ||| | ||| |||| | ||||||| |||||  ||
Db   1526 CCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCCA 1585

Qy   1854 ACAGCGCAGCAAGTGAGGCTCCC 1876
          ||| ||||| | ||||| |||
Db   1586 GGCACGGAGGAAGTGACCAGCCC 1608
```

Search completed: July 21, 2004, 12:27:13
Job time : 255 secs

```
                 GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      July 21, 2004, 03:46:00 ; Search time 1509 Seconds
                                        (without alignments)
                             11031.060 Million cell updates/sec

Title:           US-09-270-437D-6
Perfect score:   3412
Sequence:        1 ggcagcggaggaggcgagga.........aaccttgaaaatgtttattt 3412

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3191023 seqs, 2439312756 residues

Total number of hits satisfying chosen parameters:      6382046

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications NA:*
             1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
             2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
             3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
             4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
             5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
             6:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
             7:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
             8:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
             9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
            10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
            11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
            12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
            13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
            14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
            15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
            16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
            17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
            18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
            19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3411	100.0	3412	9	US-09-899-651-6	Sequence 6, Appli
2	3320.6	97.3	3694	9	US-09-764-864-329	Sequence 329, App
3	3143	92.1	3283	9	US-09-889-651-8	Sequence 8, Appli
4	3116.4	91.3	3905	10	US-09-814-353-20241	Sequence 20241, A
5	3114.8	91.3	3667	15	US-10-097-340-146	Sequence 146, App
6	3114.8	91.3	3667	17	US-10-648-593-48	Sequence 48, Appl
7	1989.8	58.3	2290	9	US-09-764-864-332	Sequence 332, App
8	1423.8	41.7	1707	16	US-10-262-445-39	Sequence 39, Appl
9	1096.8	32.1	1186	9	US-09-764-864-749	Sequence 749, App
10	776.2	22.7	822	9	US-09-764-864-330	Sequence 330, App
11	712	20.9	1740	9	US-09-735-705-347	Sequence 347, App
12	712	20.9	1740	9	US-09-850-716A-347	Sequence 347, App
13	712	20.9	1740	9	US-09-897-778-347	Sequence 347, App
14	712	20.9	1740	13	US-10-007-700-347	Sequence 347, App
15	712	20.9	1740	15	US-10-117-982-347	Sequence 347, App
16	712	20.9	1740	15	US-10-117-982-478	Sequence 478, App
17	712	20.9	1740	16	US-10-313-986-347	Sequence 347, App
18	712	20.9	1740	16	US-10-313-986-478	Sequence 478, App
19	712	20.9	1743	9	US-09-897-778-447	Sequence 447, App
20	712	20.9	1743	9	US-09-897-778-450	Sequence 450, App
21	712	20.9	1743	13	US-10-007-700-447	Sequence 447, App
22	712	20.9	1743	13	US-10-007-700-450	Sequence 450, App
23	712	20.9	1743	15	US-10-117-982-447	Sequence 447, App
24	712	20.9	1743	15	US-10-117-982-450	Sequence 450, App
25	712	20.9	1743	16	US-10-313-986-447	Sequence 447, App
26	712	20.9	1743	16	US-10-313-986-450	Sequence 450, App
27	712	20.9	1799	16	US-10-313-986-485	Sequence 485, App
28	712	20.9	34555	15	US-10-117-982-479	Sequence 479, App
29	712	20.9	34555	16	US-10-313-986-479	Sequence 479, App
30	711.8	20.9	1740	15	US-10-117-982-483	Sequence 483, App
31	711.8	20.8	1740	16	US-10-313-986-483	Sequence 483, App
32	710.4	20.8	4159	9	US-09-899-651-4	Sequence 4, Appli
33	710.4	20.8	4181	9	US-09-735-705-175	Sequence 175, App
34	710.4	20.8	4181	9	US-09-954-456-715	Sequence 715, App
35	710.4	20.8	4181	9	US-09-850-716A-175	Sequence 175, App
36	710.4	20.8	4181	9	US-09-897-778-175	Sequence 175, App
37	710.4	20.8	4181	10	US-09-466-396A-175	Sequence 175, App
38	710.4	20.8	4181	13	US-10-007-700-175	Sequence 175, App
39	710.4	20.8	4181	15	US-10-117-982-175	Sequence 175, App
40	710.4	20.8	4181	16	US-10-313-986-175	Sequence 175, App
41	710.4	20.8	4434	16	US-10-116-802-145	Sequence 145, App
42	709	20.8	1764	9	US-09-850-716A-428	Sequence 428, App
43	709	20.8	1764	9	US-09-897-778-428	Sequence 428, App
44	709	20.8	1764	13	US-10-007-700-428	Sequence 428, App
45	709	20.8	1764	15	US-10-117-982-428	Sequence 428, App

ALIGNMENTS

```
RESULT 1
US-09-899-651-6
; Sequence 6, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Gure, Ali
;  APPLICANT: Tsang, Solam
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Jager, Eike
;  APPLICANT: Knuth, Alexander
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
;  TITLE OF INVENTION: Associated
;  TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
;  LENGTH: 3412
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-899-651-6

  Query Match            100.0%;  Score 3411;  DB 9;  Length 3412;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3412;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy         1 GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGGGGAGGCGGCTCTCGGGG 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GGCAGCGGAGGAGGCGAGGAGCGCCGGGTACCGGGGGAGGCGGCTCTCGGGG 60

Qy        61 AAGAGACGGATGATGAACAAGCTTTACATCGGAACCTGAGCCCCGGGTCACCGCGCGAC 120
```

Db 1141 CAACAAGCCAATCTGATCCGAGGTTGAACCTGCAGGTTGAACCTGCAGCGCACTTGGCATCTTTCAACAGGA 1200

Qy 1201 CTGTCCGTGCTATCTCCACCACCCACCTCCGGGACTCTTCCAGCTGAGAGTCCCCGGGAGCTGCCCCTACCAC 1260
Db 1201 CTGTCCGTGCTATCTCCACCACCCACCTCCGGGACTCTTCCAGCTGAGAGTCCCCGGGAGCTGCCCCTACCAC 1260

Qy 1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCTGTACCCCCATCACCAGTTGGCCCG 1320
Db 1261 CCCTTCACTACCCACTCCGGATACTTCTCCAGCTGTACCCCCATCACCAGTTGGCCCG 1320

Qy 1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGGATTGTGAATCTCTTCATCCCAACCCAG 1380
Db 1321 TTCCCGCATCATCACTCTTATCCAGAGCAGGAGGATTGTGAATCTCTTCATCCCAACCCAG 1380

Qy 1381 GCTGTGGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACATCGGCGAGATTCGCC 1440
Db 1381 GCTGTGGGGCGCCATCATCGGGAAGAAGGGGCACACATCAAACATCGGCGAGATTCGCC 1440

Qy 1441 GGAGCCTCTATCAAGATGCCCCTGCCGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500
Db 1441 GGAGCCTCTATCAAGATGCCCCTGCCGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC 1500

Qy 1501 ATCACCCGGCCACCCGGAAGCCCAGTTCAAGGCCATATCAGAGTGCCC 1560
Db 1501 ATCACCCGGCCACCCGGAAGCCCAGTTCAAGGCCATATCAGAGTGCCC 1560

Qy 1561 GAGGAAAACTTCTTTAACCCCAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620
Db 1561 GAGGAAAACTTCTTTAACCCCAAGAAGAAGTGGAAGCGCATATCAGAGTGCCC 1620

Qy 1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGAACGGTGGCAAGACTGCGAGAAC 1680
Db 1621 TCTTCCACAGCTGGCCGGGTGATTGGCAAGAACGGTGGCAAGACTGCGAGAAC 1680

Qy 1681 TTAACCAGTGCAGAAGTCATCGTCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTG 1740
Db 1681 TTAACCAGTGCAGAAGTCATCGTCCTCGTGACCAAAGCCAGATGAAAATGAGGAAGTG 1740

Qy 1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCGGAATACCCTCAGGGAATCAGGGAA 1800
Db 1741 ATCGTCAGAATTATCGGGCACTTCTTTGCTCAGCCGGAATACCCTCAGGGAATCAGGGAA 1800

Qy 1801 ATTGTACAACAGGTGAAGCAGCAGCAGAGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860
Db 1801 ATTGTACAACAGGTGAAGCAGCAGCAGAGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGC 1860

Qy 1861 AGCAAGTGAGGCTCCCACAGGACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920
Db 1861 AGCAAGTGAGGCTCCCACAGGACCAGCAAAACAACGGATGAATGTAGCCCTTCCAACAC 1920

Qy 1921 CTGACAGAATGAGACCAGATGACAAACCAGCGAGCAAACCAAAGACCATCTGAG 1980
Db 1921 CTGACAGAATGAGACCAGATGACAAACCAGCGAGCAAACCAAAGACCATCTGAG 1980

Qy 1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGAGTCTGCCTCACAGCGC 2040
Db 1981 GAATGAGAAGTCTGCGGAGGCGGCCAGGGACTCTGCCGAGGAGTCTGCCTCACAGGGGC 2040

Qy 2041 CGGAGGGGCCTTCTGCGAGGCTTCAGCGGAAGTTGCCAGAACCACCAGAGCCCCTTCCCGCC 2100
Db 2041 CGGAGGGGCCTTCTGCGAGGCTTCAGCGGAAGTTGCCAGAACCACCAGAGCCCCTTCCCGCC 2100

Qy 2101 CCCCAGGGCTTCTGCGAGGCTTCAGCGGATCTCACCACTCACTCCGGATCTCTCCGTTGAA 2160
Db 2101 CCCCAGGGCTTCTGCGAGGCTTCAGCGGATCTCACCACTCACTCCGGATCTCTCCGTTGAA 2160

Qy 2161 CTCCCACGACGCTATCCCTTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGC 2220
Db 2161 CTCCCACGACGCTATCCCTTTAGTTGAACATAGGTGAACGTGTTCAAAGCCAAGC 2220

Qy 2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280
Db 2221 AAAATGCACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAA 2280

Db 61 AAGAGACGGATGATGAACAAGCTTACATCGGGAACACTGGAGCCCCGTCACCGCCGAC 120

Qy 121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGCCGGGGACCAGTCCTGCTGAAG 180
Db 121 GACCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGCCGGGGACCAGTCCTGCTGAAG 180

Qy 181 TCCGGCTACGGCTACGGCCTTCGTGGGACTACCCCGACCAGGAATGGGCCATCGAGACC 240
Db 181 TCCGGCTACGGCTACGGCCTTCGTGGGACTACCCCGACCAGGAATGGGCCATCGAGACC 240

Qy 241 CTCTCCGGGTAAAGTGGAATTCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAA 300
Db 241 CTCTCCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAA 300

Qy 301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTTCCTCACCTGCAGTGGGAGGTG 360
Db 301 AAGCTAAGGAGCAGGAAAATTCAGATTCGAAACATCCCTTCCTCACCTGCAGTGGGAGGTG 360

Qy 361 TTGGATGGACTTTTGGCTCAATATGGCACAGTGGGACAAGTCAACACAGAC 420
Db 361 TTGGATGGACTTTTGGCTCAATATGGCACAGTGGGACAAGTCAACACAGAC 420

Qy 421 ACAGAAACCGCCGTGTCACAACGTCACATATGCAACAAGAGAAGCAAAATAGCCATG 480
Db 421 ACAGAAACCGCCGTGTCACAACGTCACATATGCAACAAGAGAAGCAAAATAGCCATG 480

Qy 481 GAGAAGCTAAGGCGGCATCAGTTGACATCACTCCTTCCTTCAAGATTCCTACATCCCGAT 540
Db 481 GAGAAGCTAAGGCGGCATCAGTTGACATCACTCCTTCCTTCAAGATTCCTACATCCCGAT 540

Qy 541 GAAGAGGTGAGCTCCCCTTCGCCCCTCGAGCCCAGCTGGGACCACTCTTCCCGG 600
Db 541 GAAGAGGTGAGCTCCCCTTCGCCCCTCGAGCCCAGCTGGGACCACTCTTCCCGG 600

Qy 601 GCGCAAGGGCCACGCCCCTGGGGACACTTCTCAGCGCACACAGATTGATTCCCGCTGCGG 660
Db 601 GCGCAAGGGCCACGCCCCTGGGGACACTTCTCAGCGCACACAGATTGATTCCCGCTGCGG 660

Qy 661 ATCGGTCTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAAGGAGGGGCTTGACCATAAAG 720
Db 661 ATCCTGGTCTCCCCACCCAGTTTGTTGGTGCCATCATCGGGAAAAGGAGGGGCTTGACCATAAAG 720

Qy 721 AACATCACTCAAGCAGACGAACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
Db 721 AACATCACTCAAGCAGACGAACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780

Qy 781 GCGAGGAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCCATGCCGGCATG 840
Db 781 GCGAGGAAGCCTGTCACCATGCCACCCCAGAGGGGACTTCTGAAGCATGCCGGCATG 840

Qy 841 ATTCTTGAAATCATGCAGAAATAGCCGAGAAGCCAAAACTACCGCCGAAGAGATTCCTCTG 900
Db 841 ATTCTTGAAATCATGCAGAAATAGCCGAGAAGCCAAAACTACCGCCGAAGAGATTCCTCTG 900

Qy 901 AAAATCTTGGCACACACAATGGCTTGGTTGGAAGACTGATTGGAAGACGGCAGAAATTTG 960
Db 901 AAAATCTTGCACACACAATGGCTTGGTTGGAAGACTGATTGGAAGACGGCAGAAATTTG 960

Qy 961 AAGAAAAATTGAACATGAAACATGAAACACCAGGGACCAAGAACAACATCACTGTGAAGC 1020
Db 961 AAGAAAAATTGAACATGAAACATGAAACACCAGGGACCAAGAACAACATCACTGTGAAGC 1020

Qy 1021 ATATACACACCCGAAAGAACATCACTGTGAAAGCCTGTGAGGCCTGTGCCAGTGCT 1080
Db 1021 ATATACACACCCGAAAGAACATCACTGTGAAAGCCTGTGAGGCCTGTGCCAGTGCT 1080

Qy 1081 GAGATAGAGATTATGAAGAGTCTGGCGTGAAAATGATATGCTGCGTGTTAAC 1140
Db 1081 GAGATAGAGATTATGAAGAGTCTGGCGTGAAAATGATATGCTGCGTGTTAAC 1140

Qy 1141 CAACAAGCCAATCTGATCCGAGGTTGAACCTGCAGGTTCAACAGGA 1200

```
Qy   3361 AGCCAAGAACCNATATGCCTTCTTCTTTTGGACAAACCTTGAAAAATGTTTATTT 3412
          ||||||||||| ||||||||||||||||||||||||||| |||||||||||||||
Db   3361 AGCCAAGAACCNATATGCCTTCTTCTTTTGGACAAACCCTTGAAAAATGTTTATTT 3412
```

RESULT 2
US-09-764-864-329
; Sequence 329, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 329
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-329

Query Match   97.3%;   Score 3320.6;   DB 9;   Length 3694;
Best Local Similarity   99.8%;   Pred. No. 0;
Matches 3343;   Conservative 1;   Mismatches 1;   Indels 3;   Gaps 2;

```
Qy    68 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGACCTCACCGGCCGACACCTCC 127
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91 GGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGACCTCACCGGCCGACACCTCC 150

Qy   128 GGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGCGGGGACAGGTCCTGCTGAAGTCCGGCT 187
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 GGCAGCTCTTTGGGGACAGGAAGCTGCCCCTGGGCGGGGACAGGTCCTGCTGAAGTCCGGCT 210

Qy   188 ACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGCGCCGCCATCCAGAGACCCCTCCG 247
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ACGCCTTCGTGGACTACCCCGACCAGAACTGGGCCATCGCGCCGCCATCCAGAGACCCCTCCG 270

Qy   248 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAA 307
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 GTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAA 330

Qy   308 GGAGCAGGAGAAATTCAGATTGAAACATCCCTCCTCAGTGGAGGTGTTGGATG 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 GGAGCAGGAGAAATTCAGATTGAAACATCCCTCCTCAGTGGAGGTGTTGGATG 390

Qy   368 GACTTTTGGCTCAATATGGACAGTGGACAGTGGAACAAGTCAACACAGAACAGAAA 427
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 GACT-TTGGCTCAATATGGACAGTGGACAGTGGAACAAGTCAACACAGAACAGAAA 450

Qy   428 CCGCCGTTGTCAAGTCACATATGCAACAGAGAAGAGCAAAAATAGCCATGGAGAAGC 487
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 CCGCCGTTGTCAAGTCACATATGCAACAGAGAAGAGCAAAAATAGCCATGGAGAAGC 510

Qy   488 TAAGCGGGCATCAGTTTGAGACTACTCCTTCAGATTTCCTACATCCCGATGAAGAGAG 547
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511 TAAGCGGGCATCAGTTTGAGACTACTCCTTCAGATTTCCTACATCCCGATGAAGAGAG 570

Qy   548 TGAGCTCCCCTTCGCCCCCTCAGCGGAGCCCAGCGCGTGGGAGCAG 607
         |||||||||||||||||||||||||||||||||||||||||||||
Db   571 TGAGCTCCCCTTCGCCCCCTCAGCGGAGCCCAGCGCGTGGGAGCAG 630

Qy   608 GCCACCCCAGTTCTTCAGGCCGCAGACAGATTGATTTCCCGCTCGCGGATCCTGG 667
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631 GCCACCCCAGTTCTTCAGGCCGCAGACAGATTGATTTCCCGCTCGCGGATCCTGG 690

Qy   668 TCCCCACCCAGTTGTTGGTGCCATCGCGAAAGGAGGGCTTGACCATCGGAAAGAACATCA 727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   691 TCCCCACCCAGTTGTTGGTGCCATCGCGAAAGGAGGGCTTGACCATCGGAAAGAACATCA 750

Qy   728 CTAAGCAGACACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGAGCTGCGAGAGA 787
```

```
Qy  2281 GGGAGAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGGTAATATAT 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GGGAGAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCGGTAATATAT 2340

Qy  2341 TTTAGAAATAATATCAAATAATATCAATAATTTTAATCAATTATTAATCAATTATTAATTTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 TTTAGAAATAATATCAAATAATATCAATAATTTTAATCAATTATTAATCAATTATTAATTTT 2400

Qy  2401 TTTTCTTTTAAAGAGAAGAAGCAGCGCTTTCTAGACTTTTAAAGAATAAAGTCTTTGGGAG 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTTCTTTTAAAGAGAAGAAGCAGCGCTTTCTAGACTTTTAAAGAATAAAGTCTTTGGGAG 2460

Qy  2461 GTCTCACGGTGTGAGAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCTCACGGTGTGAGAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA 2520

Qy  2521 CTCGTCGGAAGGAAAACTCTGCCACTCGTGCCTGTCTAGCTCATGCAACACCCATT 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTGTATGTCAACAGAAGGGATA 2580

Qy  2581 CCGTCTCCCTGAAGAGGAAAACTCTGTCACTCTGTCATCCTGTGTCTAGCTCATCACACCCATT 2640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCGTCTCCCTGAAGAGGAAAACTCTGTCACTCTGTCATCCTGTGTCTAGCTCATCACACCCATT 2640

Qy  2641 TCTCTTGCTTCACAGGTTTTAAAACTGGTTTTTTTGCATATGCTATATAATTCCTGTCT 2700
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TCTCTTGCTTCACAGGTTTTAAAACTGGTTTTTTTGCATATGCTATATAATTCCTGTCT 2700

Qy  2701 CTCTCGTTTATCTCTCCCCTTCCCCCTTCTCTCCATCTCCATTCTCTTTTGAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTCTCGTTTATCTCTCCCCTTCCCCCTTCTCTCCATCTCCATTCTCTTTTGAA 2760

Qy  2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGACACCCCCCCCCCCCCAGGCAAAGCA 2820
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TTTCCTCATCCTCCATCTCAATCCGTATCTACGGACACCCCCCCCCCCCCCAGGCAAAGCA 2820

Qy  2821 GTGCTCTGAGTATCACATCACACACAAAGGAACACAAAGCGAAACACACAGAACCAGCCTCAA 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GTGCTCTGAGTATCACATCACACACAAAGGAACACAAAGCGAAACACACAGAACCAGCCTCAA 2880

Qy  2881 CTTACACTTGGTTACTTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTTACACTTGGTTACTTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGGGTTTTGGAAG 2940

Qy  2941 AGGAAAAACAGGAACCCACCACCAACCAATCAACCAATCAACCAATCAACCAATTCCACAATG 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 AGGAAAAACAGGAACCCACCACCAACCAATCAACCAATCAACCAATTCCACAATG 3000

Qy  3001 AAAGAATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 AAAGAATGTATTTTGTCTTTTTGGTGTATAAGCCATCAATATTCAGCAAAATG 3060

Qy  3061 ATTCCTTTCTTTAAAAAAAAATGTGGAAGTTAGAAATTTACCAAGGTTGTTGGCC 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 ATTCCTTTCTTTAAAAAAAAATGTGGAAGTTAGAAATTTACCAAGGTTGTTGGCC 3120

Qy  3121 CAGGGGCGTTAAATTCACAGATTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT 3180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 CAGGGGCGTTAAATTCACAGATTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACT 3180

Qy  3181 GTGTTTTTACCTTGACACCTTGCTCTTTTGTTCTTGTGTTCCTTAGAGATTTTGTAAAGCTTGCATGATAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 GTGTTTTTACCTTGACACCTTGCTCTTTTGTTCTTGTGTTCCTTAGAGATTTTGTAAAGCTTGCATGATAG 3240

Qy  3241 TTGGAGCATTTTTTATTTTTTTTATTTTTTTTTATAAAAATGAGTTGGAAAAAATAAGATATCAACT 3300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TTGGAGCATTTTTTATTTTTTTTATTTTTTTTATAAAAATGAGTTGGAAAAAATAAGATATCAACT 3300

Qy  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTGTCAACAGCTGTGTCTTCCGCT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 GCCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTGTCAACAGCTGTGTCTTCCGCT 3360
```

```
Db  1831 AACAGGTGAAGCAGGAGCAGGAGCAGAAATACCCTCAGGGAGTCGGCCTCACAGCGCAGCAAGT 1890
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   751 CTAAGCAGACCCGGGTGTAGATATCCATAGAAAGAGAACTCTGGAGCTGCGAGAGA 810

Qy  1868 GAGGCTCCCACAGGCACCAGCCACCGATGAATGTAGCCCTTCCAACACCTGACAG 1927
Db   788 AGCCTGTCACCATCCATGCCACCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 847

Db  1891 GAGGCTCCCACAGGCACCAGCCACCGATGAATGTAGCCCTTCCAACACCTGACAG 1950
Qy   811 AGCCTGTCACCATCCATGCCACCCCCAGAGGGGACTTCTGAAGCATGCCGCATGATTCTTG 870

Qy  1928 AATGAGACCAAAGCGACCCAGATGGGGAGCAAACCAATCTGAGGGAATGAG 1987
Db   848 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 907

Db  1951 AATGAGACCAAAGCGACCCAGATGGGGAGCAAACCAATCTGAGGGAATGAG 2010
Qy   871 AAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCT 930

Qy  1988 AAGTCTGCGGAGGCGCGGAGGCTGCGAGGCCCTGAGAACCCCGAGGGCGCCGAGGAG 2047
Db   908 TGGCACCACAAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAAA 967

Db  2011 AAGTCTGCGGAGGCGCGGAGGCTGCGAGGCCCTGAGAACCCCGAGGGCGCCGAGGAG 2070
Qy   931 TGGCACCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGCAGAAATTTGAAGAAA 990

Qy  2048 GGGCGGGGAAGTTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTTCCCGCCCCCCAGG 2107
Db   968 TTGAAACATGAAACAGGAGCACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1027

Db  2071 GGGCGGGGAAGTTCAGCCAGGTTTGCCAGAACCACCGAGCCCCGCCTTCCCGCCCCCCAGG 2130
Qy   991 TTGAAACATGAAACAGGAGCACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCATATACA 1050

Qy  2108 GCTTCTGCAGGCTTCAGCCATCCACTCACCTCGGATCTCTCCTGAACTCCCAC 2167
Db  1028 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTCGTCCAGTGCTGAGATAG 1087

Db  2131 GCTTCTGCAGGCTTCAGCCATCCACTCACCTCGGATCTCTCCTGAACTCCCAC 2190
Qy  1051 ACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTGTCGTCCAGTGCTGAGATAG 1110

Qy  2168 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2227
Db  1088 AGATTATGAAGAAGTCGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAG 1147

Db  2191 GACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAAAATGC 2250
Qy  1111 AGATTATGAAGAAGTCGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCAACAG 1170

Qy  2228 ACACCCTTTTTCTGTGGCAAATCGTCTCTGATACATGTGTACATATTAGAAGGGAAGA 2287
Db  1148 CCAATCTGATCCCAGGTTGAACCTTGGCACTCTTTTCAACAGGACTGTCCG 1207

Db  2251 ACACCCTTTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAGGGAAGA 2310
Qy  1171 CCAATCTGATCCCAGGTTGAACCTTGGCACTCTTTTCAACAGGACTGTCCG 1230

Qy  2288 TGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCTGCAGCGGTAATATATTTTAGAA 2347
Db  1208 TGCTATCTCCACCAGCAGGAGCTCCCGCGGAGCTCCCCCGCTGCCCCCTACCACCCCTTCA 1267

Db  2311 TGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCTGCAGCGGTAATATATTTTAGAA 2370
Qy  1231 TGCTATCTCCACCAGCAGGAGCTCCCGCGGAGCTCCCCCGCTGCCCCCTACCACCCCTTCA 1290

Qy  2348 ATAATATATCAAATAACTCAACTAACTCCAATTTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2407
Db  1268 CTACCCACTCCGGATACTTCTCCAGCCTGACCCTGTACCCCATCACCACCGTTCCCGC 1327

Db  2371 ATATATATCAAATAACTCAACTAACTCCAATTTTTAAAGAATAAAGTCTTTGGGAGGTCTCAC 2430
Qy  1291 CTACCCACTCCGGATACTTCTCCAGCCTGACCCTGTACCCCATCACCACCGTTCCCGC 1350

Qy  2408 TTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGG 2467
Db  1328 ATCATCACTCTTATCCAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGG 1387

Db  2431 TTTTAAAGAGAAGCAGGCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGTCTGTGG 2490
Qy  1351 ATCATCACTCTTATCCAGACGAGGAGATTGTGAATCTCTTCATCCCAACCCAGGCTGTGG 1410

Qy  2468 GGTGTAGAGAGAGCTTGGAGCTTTGAGGCCACACCGACAACAAATTCAACGAGGGGAAATCTCGTCG 2527
Db  1388 GCGCACCCGAAGCCCCTGCGGAAGCCCCAGAGCCCCAGGCCCACATCATCACCG 1447

Db  2491 GGTGTAGAGAGAGCTTGGAGCTTTGAGGCCACACCGACAACAAATTCAACGAGGGGAAATCTCGTCG 2550
Qy  1411 GCGCACCCGAAGCCCCTGCGGAAGCCCCAGAGCCCCAGGCCCACATCATCACCG 1470

Qy  2528 GAAGGCACTCACGGCAGTTCGGATCACCTGTGATGTCAACAGAAGGGATACCGTCTC 2587
Db  1448 CTATCAAGAGATTGCCCTGCCCAGTTCAAGGCCCCAGGCCCAGAGATGGTCATCATCACCG 1507

Db  2551 GAAGGCACTCACGGCAGTTCGGATCACCTGTGATGTCAACAGAAGGGATACCGTCTC 2610
Qy  1471 CTATCAAGAGATTGCCCTGCCCAGTTCAAGGCCCCAGGCCCAGAGATGGTCATCATCACCG 1530

Qy  2588 CTTGAAGGAGAAACTCACGGTTTTTGCATCGGTTTTTGCATACTGTACTGTCTAGTGTATATAATTCTCTTT 2647
Db  1508 GGCACCCGAAGCCCCAGTTCAAGGCCCAGGACCGATCTTTGGGAAACTGAAAAGAGGAAA 1567

Db  2611 CTTGAAGGAGAAACTCTGCACTCCTGTCACTCCCCATGCCATGCCCATGCGCGTCTCTAGTGTATATAATTCTCTTT 2670
Qy  1531 GGCCACCCGAAGCCCCAGTTCAAGGCCCAGGACCGATCTTTGGGAAACTGAAAAGAGGAAA 1590

Qy  2648 GCTTCACAGGTTTTAAACTGGTTTTTGCATCTGGTTTTTGCATAATAATTCTCTGTCTCTCTG 2707
Db  1568 ACTTCTTTAAACCCAAAGAAGAGAGTGAAGCTGAAATGAGGGAAGTGCCCTCTTCCA 1627

Db  2671 GCTTCACAGGTTTTAAACTGGTTTTTGCATCTGGTTTTTGCATAATAATTCTCTGTCTCTCTG 2730
Qy  1591 ACTTCTTTAAACCCAAAGAAGAGAGTGAAGCTGAAASGCGCATATCAGAGTGCCCTCTTCCA 1650

Qy  2708 TTTATCTCTCCCCATCTCAATCCCGTATCCGTATCTCACGCACCCCCCCCCCCATTCTCTTT 2767
Db  1628 CAGCTGGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTCAGACTTGTAC 1687

Db  2731 TTTATCTCTCCCCATCTCAATCCCGTATCCGTATCTCACGCACCCCCCCCCCCATTCTCTTT 2790
Qy  1651 CAGCTGGCCGGGTGATTGGCAAAGTGGCAAGACCGTGAACGAACTTAACCA 1710

Qy  2768 ATCCCTCCATCAATCCGTATCTACGCA--CCCCCCCCCCCCCCCAGGCAAAGCAGTGCT 2825
Db  1688 GTGCAGAAGTCATCGTGCCTCGTGACCAAAGACGGCAGATTGACGTGTCA 1747

Db  2791 ATCCCTCCATCAATCCGTATCTACGCACCCCCCCCCCCCCCCAGGCAAAGCAGTGCT 2850
Qy  1711 GTGCAGAAGTCATCGTGCCTCGTGACCAAAGGCGACTGAGGAAGTGATCGTCA 1770

Qy  2826 CTGAGTATCACATCACAAACAGAACAAAAGCAAACACAACAAACCAGCCTCAACTTAC 2885
Db  1748 GAATTATCGGGCACTTCTTTGCTCAGCGACTTGCACACGGCAAGATCAGCGGAAATTGTAC 1807

Db  2851 CTGAGTATCACATCACATCACAAAAGGAACAAAAGGAACACAACAAACCAGCCTCAACTTAC 2910
Qy  1771 GAATTATCGGGCACTTCTTTGCTCAGCGACTTGCACACGGCAAGATCAGCGGAAATTGTAC 1830

Qy  2886 ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAGAGGAA 2945
Db  1808 AACAGGTGAAGCAGGAGCAGGAGCAGAAATACCCTCAGGGAGTCGGCCTCACAGCGCCAGCAAGT 1867

Db  2911 ACTTGGTTACTCAAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAGAGGAA 2970
```

```
Qy  121 GACTTCCGCCAGCTCTTTGGGACCAGCAGAGAAGCTGCCCTGCGCGGGACCAGGTCCTGCTGAAG 180
        ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db  121 GACTTCCGCCAGCTCTTTGGGACCAGCAGAGAAGCTGCCCTGCTGGGGACCAGGTCCTGCTGAAG 180

Qy  181 TCCGGCTACGCCTTCGTGACTACCACCCGACCAGAGATCGGGGCCATCGGCCGCGCCATCGAGACC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TCCGGCTACGCCTTCGTGACTACCACCCGACCAGAGATCGGGGCCATCGGCCGCGCCATCGAGACC 240

Qy  241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAA 300

Qy  301 AAGCTAAGGAGCAGGAAAATTCAGATTCGGAAACATCCCTCCTGCACTGCAGTGGGAGGTG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AAGCTAAGGAGCAGGAAAATTCAGATTCGGAAACATCCCTCCTGCACTGCAGTGGGAGGTG 360

Qy  361 TTTGGATGGACTTTTGGCTCAATATGGGACAGTGGAACAGTGATAGTTCG 420
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTTGGATGGACTTTTGGCTCAATATGGGACAGTGGAACAGTGATAGTTCG 420

Qy  421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAATAGCCATG 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ACAGAAACCGCCGTTGTCAACGTCACATATGCAACAGAGAAGCAAAATAGCCATG 480

Qy  481 GAGAAGCTAAGCGGCGCATCAGTTGAGAACTACTCCTTCAAGATTTCTTCCGCTAGCC 540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGAAGCTAAGCGGCGCATCAGTTGAGAACTACTCCTTCAAGATTTCTTCCGCTAGCC 540

Qy  541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGGACCACTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAAGAGGTGAGCTCCCCTTCGCCCCTGGGGGACCACTCTCAGGCCAGACAGATTGATTTCCCGCTGCGG 600

Qy  601 GAGCAAGGCCACGCCCAGCCCCGG 660
        ||||||||||||||||||||||||
Db  601 GAGCAAGGCCACGCCCAGCCCCGG 660

Qy  661 ATCCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAAGAGGGCTTGACCATAAAG 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 ATCCCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAAAGAGGGCTTGACCATAAAG 720

Qy  721 AACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AACATCACTAAGCACAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAGAACTCTGGAGCT 780

Qy  781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GCAGAGAAGCCTGTCACCATCCATGCCACCCCAGAGGGACTTCTGAAGCATGCCGCATG 840

Qy  841 ATTCTGAAATCATGCAGAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 ATTCTGAAATCATGCAGAGAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTG 900

Qy  901 AAAATCTTGGCACACAATGCTTGGTTGGAAGACTGATTGGAAGAGAGGCAGAAATTTG 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 AAAATCTTGGCACACAATGCTTGGTTGGAAGACTGATTGGAAGAGAGGCAGAAATTTG 960

Qy  961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 AAGAAAATTGAACATGAAACAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGC 1020

Qy 1021 ATATCAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTCCCAGTGCT 1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 ATATCAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCCTTGTCCCAGTGCT 1080

Qy 1081 GAGATAGAGATTATGAAGAAGCTGCGTGAAATGATATGTGGCTGTTAAC 1140
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GAGATAGAGATTATGAAGAAGCTGCGTGAAATGATATGTGGCTGTTAAC 1140

Qy 1141 CAACAAGCCAATCTGATCCCAGGTGTGAACCTCAGCGCCACTTGGCCATCTTTTCAACAGGA 1200
        ||||||
Db 1141 --------------------------------------------------------------- 1140
```

```
Qy 2946 AACAGGAACCCACCCAAACCAACCAATCAACCAAACAAAGAAAAAATTCCACAATGAAAGA 3005
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2971 AACAGAGAACCCACCCAAACCAACCAACCAATCAAACAAACAAAGAAAAAATTCCACAATGAAAGA 3030

Qy 3006 ATGTATTTGTCTTTTGCATTTGTGTGATAAAGCCATCATCAAATATTCAGCAAAATGATTCC 3065
        ||||||||||||||||||||||||||||||||||||||||||||
Db 3031 ATGTATTTGTCTTTTGCATTTGTGTATAAAGCCATCATCAAATATTCAGCAAAATGATTCC 3090

Qy 3066 TTTCTTT-AAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG 3124
        ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3091 TTTCTTTAAAAAAAAAAAATGTGGAGGAAAGTAGAAATTTACCAAGGTTGTTGGCCCAGG 3150

Qy 3125 GCGTTAAATTCACAGATTTTTTTAACGAGAAAAACACACGAGAAGAAGCTACCTCAGGTGT 3184
        |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db 3151 GCGTTAAATTCACAGATTTTTTTTACGAGAAAAACACACGAGAAGAAGCTACCTCAGGTGT 3210

Qy 3185 TTTTACCTCAGCACCTTGCTCCTTGTGTTTCCCTTTAGAGATTTTGTAAAGCTGATAGTTCG 3244
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3211 TTTTACCTCAGCACCTTGCTCCTTGTGTTTCCCTTTAGAGATTTTGTAAAGCTGATAGTTCG 3270

Qy 3245 AGCATTTTTTTATTTTTTAATAAAATGAGTTGGAAAAAAAATAAGATATCAACTGCCA 3304
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3271 AGCATTTTTTTATTTTTTTTAATAAAATGAGTTGGAAAAAAAAATAAGATATCAACTGCCA 3330

Qy 3305 GCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGTGTTCGAATTGTCTTCCGCTAGCC 3364
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3331 GCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGTGTTCGAATTGTCTTCCGCTAGCC 3390

Qy 3365 AAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3412
        |||||||  ||||||||||||||||||||||||||||||||||||||||
Db 3391 AAGAACCTATATGGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3438
```

RESULT 3
US-09-899-651-8
; Sequence 8, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Eike
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-8

Query Match 92.1%; Score 3143; DB 9; Length 3283;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 3283; Conservative 0; Mismatches 0; Indels 129; Gaps 1;

```
Qy    1 GGCAGCCGGAGGAGGCGGAGGAGGCCGGTACCGGGGGAGCCCGGGGGCCTCTCGGGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCAGCCGGAGGAGGCGGAGGAGGCCGGTACCGGGGGAGCCCGGGGGCCTCTCGGGG 60

Qy   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGGAC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGAGACGGATGATGAACAAGCTTTACATCGGGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGGAC 120
```

```
Qy  1201  CTGTCCGTGCTATCTCCACCAGCAGGGCCCCGCGGAGCTCCCCGCTGCCCCCTACCAC  1260
          ||||||||||||||||||||||||||||||||||||||||||||
Db  1141  ------------------------------------------------------------  1140

Qy  1261  CCCTTCACTACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG  1320
          ||||||||||||||||||||||||||||||||||
Db  1141  -------ACCCACTCCGGATACTTCTCCAGCCTGTACCCCCATCACCAGTTTGGCCCG   1191

Qy  1321  TTCCCGCATCATCACTCTTATCCGAGACAGGAGATTGTGAATCTTCATCCCAACCCAG  1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1192  TTCCCGCATCATCACTCTTATCCGAGACAGGAGATTGTGAATCTTCATCCCAACCCAG  1251

Qy  1381  GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCC  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1252  GCTGTGGGCGCCATCATCGGGAAGAAGAGGGGCCACACATCAAACAGCTGGCGAGATTCGCC  1311

Qy  1441  GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1312  GGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATC  1371

Qy  1501  ATCACCGGGCACCGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAA   1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1372  ATCACCGGGCACCGAAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAA   1431

Qy  1561  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGCAAGACCGTGAACGAACTGCCC  1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1432  GAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGCAAGACCGTGAACGAACTGCCC  1491

Qy  1621  TCTTCCACAGTGTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTCGAGAC  1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1492  TCTTCCACAGTGTGGCCGGGTGATTGGCAAAGGTGGCAAGACCGTGAACGAACTCGAGAC  1551

Qy  1681  TTAACCAGTGCAGAGTCATCGTGCCTCGTGACCAAAACGCCGCAAGATGAAAATGAGGGAAGTG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1552  TTAACCAGTGCAGAGTCATCGTGCCTCGTGACCAAAACGCCGCAAGATGAAAATGAGGGAAGTG  1611

Qy  1741  ATCGTCAGATTATGGGCACTTCTTGCTAGCCAGAATACCCTCAGGGGAGTCGCCTCACAGCGC  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1612  ATCGTCAGATTATGGGCACTTCTTGCTAGCCAGAATACCCTCAGGGGAGTCGCCTCACAGCGC  1731

Qy  1801  ATTGTACACAGGTGAAGCAGCAGGAGAGGAGCAGAAATACCCTCAGGGCAAAACGGATGAATGTAGCC   1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1672  ATTGTACACAGGTGAAGCAGCAGGAGAGGAGCAGAAATACCCTCAGGGCAAAACGGATGAATGTAGCC   1731

Qy  1861  AGCAAGTGAGGCTCCCACAGGCACCAGGCAGCAGCACCATCCATTCACCTCGGATCTCTCCACAC  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1732  AGCAAGTGAGGCTCCCACAGGCACCAGGCAGCAGCACCATCCATTCACCTCGGATCTCTCCACAC  1791

Qy  1921  CTGACAGAAATGAGACCAAAACGCAGCCAGCAGCAAAGACCATCTCGAG  1980
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1792  CTGACAGAAATGAGACCAAAACGCAGCCAGCAGCAAAGACCATCTCGAG  1851

Qy  1981  GAATGAGAAGTCTGCGGAGGCGGCGGCCAGGGACTCTGCCGAGGCCCCTGAGAACCCCAGGGGGC  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1852  GAATGAGAAGTCTGCGGAGGCGGCGGCCAGGGACTCTGCCGAGGCCCCTGAGAACCCCAGGGGGC  1911

Qy  2041  CGAGGAGGGGCGGGAAGGTCAGCCAGGTTGCCAGAGAACCAACACCACCGGCCCGCTTCCCGCC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1912  CGAGGAGGGGCGGGAAGGTCAGCCAGGTTGCCAGAGAACCAACACCACCGGCCCGCTTCCCGCC  1971

Qy  2101  CCCCAGGGGCTTCTGCGAGCTTCAGCGATCCATCACTTCACCTCGGATCTCTCCGATCCTTCCTGAA  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1972  CCCCAGGGGCTTCTGCGAGCTTCAGCGATCCATCACTTCACCTCGGATCTCTCCGATCCTTCCTGAA  2031

Qy  2161  CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACTGTCAAGCCAAGCC  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2032  CTCCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACTGTCAAGCCAAGCC  2091

Qy  2221  AAAATGCACACCCCTTTTCTTGTGCGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2092  AAAATGCACACCCCTTTTCTTGTGCGCAAATCGTCTCTGTACATGTGTACATATTAGAAA  2151

Qy  2281  GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGGTGCCTGCAGCCGGTAAATATAT  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2152  GGGAAGATGTTAAGATATGTGCCTGTGGGTTACACAGGGTGCCTGCAGCCGGTAAATATAT  2211

Qy  2341  TTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2212  TTTAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT  2271

Qy  2401  TTTTCTTTTTAAAGAGAAGAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2272  TTTTCTTTTTAAAGAGAAGAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAG  2331

Qy  2461  GTCTCACGGTGTAGAGAGAGAGCTTTGAGGCCACCCGGCACAAAATTCACCCAGAGGGAAAT  2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2332  GTCTCACGGTGTAGAGAGAGAGCTTTGAGGCCACCCGGCACAAAATTCACCCAGAGGGAAAT  2391

Qy  2521  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACGAGAAGGGATA  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2392  CTCGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACGAGAAGGGATA  2451

Qy  2581  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTACACCCATT  2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2452  CCGTCTCCTTGAAGAGGAAACTCTGTCACTCCTCATGCCTGTCTAGCTACACCCATT  2511

Qy  2641  TCTCTTTGCTTCACTCTCCCTCCCCCCCGTATACTGCTATATAATTCTCATTCTTTGTCT  2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2512  TCTCTTTTGCTTCACGGTTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCATTCTTTGTCT  2571

Qy  2701  CCCTCTGTTTATCTTCCCCTCCCCTCCCCCCCCCGTATCCCGTATCCCATCTCCTTCTTCTCTCCATCTCCTTCTTCTCTCCATC  2760
          ||||||||||||||||||||||||||||||||||||||||||||
Db  2572  CTCTCGTTTATCTTCCCCTCCCCTCCCCCCCCCGTATCCCGTATCCCATCTCCTTCTTCTCTCCATC  2631

Qy  2761  TTTCCTCATCCCTCCATCTCAATCCCGTATCCCGTATCACGCACAAAAGCGAAAGCACACACCAGCCTCAA  2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2632  TTTCCTCATCCCTCCATCTCAATCCCGTATCCCGTATCACGCACAAAAGCGAAAGCACACACCAGCCTCAA  2691

Qy  2821  GTGCTCTGAGTATCACATCACACAAAAGGAACAAGAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2692  GTGCTCTGAGTATCACATCACACAAAAGGAACAAGAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAG  2751

Qy  2881  CTTACACTTGGTTACTCAAAAGAACCAACCAATCAACAAAGAAAAAAATTCCACAATG  2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2752  CTTACACTTGGTTACTCAAAAGAACCAACCAATCAACAAAGAAAAAAATTCCACAATG  2811

Qy  2941  AGGAAAACAGGAACCCACCAACCAACCAACAAAGAAAAAAAATTCCACAATG  3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2812  AGGAAAACAGGAACCCACCAACCAACCAACAAAGAAAAAAAATTCCACAATG  2871

Qy  3001  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2872  AAAGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATG  2931

Qy  3061  ATTCCTTTCTTTAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC  3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2932  ATTCCTTTCTTTAAAAAAAAAATGTGGAGAAAGTAGAAATTTACCAAGGTTGTTGGCC  2991

Qy  3121  CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG  3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2992  CAGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAG  3051

Qy  3181  GTGTTTTACCTCAGCACCTTGCTCCGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3052  GTGTTTTACCTCAGCACCTTGCTCCGTGTTTCCCTTAGAGATTTTGTAAAGCTGATAG  3111

Qy  3241  TTGGAGCATTTTTTATTTTTTAATAGAAAAATGAGTTGGAAAAAATAAGAATCAACT  3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3112  TTGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAATCAACT  3171

Qy  3301  GCCAGCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT  3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3172  GCCAGCCTGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCT  3231

Qy  3361  AGCCAAGAACCNATATGGCCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT  3412
          ||||||||||||||||
```

```
Db      3232 AGCCAAGAACCNATATGGCCTTCTTTTGGACAAACCTTGAAAATGTTTATTT 3283

RESULT 4
US-09-814-353-20241
; Sequence 20241, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT: Thompson, Pamela
;  APPLICANT: Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 20241
;    LENGTH: 3905
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: 1, 3897, 3898, 3899, 3900, 3901, 3902, 3903, 3904, 3905
;    OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20241

Query Match           91.3%;  Score 3116.4;  DB 10;  Length 3905;
Best Local Similarity 96.1%;  Pred. No. 0;
Matches 3278; Conservative 0; Mismatches 2; Indels 131; Gaps 3;
```

```
Qy     423 AGAAACCGCCGTGTCAACGTCACATATGCACAGAGAGAGCAAAATAGCCATGGA 482
           ||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db     789 AGAAACCGCCGTGTCAACGTCACATATGCAACAGAGAGAGCAAAAATAGCCATGGA 848

Qy     483 GAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCGATGA 542
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     849 GAAGCTAAGCGGGCATCAGTTGAGAACTACTCCTTCAAGATTCCTACATCCGATGA 908

Qy     543 AGAGGTGAGTCCCCTTCGCCCCTCGCCGGTGGGACCACTCTTCCCGGGA 602
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     909 AGAGGTGAGTCCCCTTCGCCCCTCGCCGGTGGGACCCACTCTTCCCGGGA 968

Qy     603 GCAAGGCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGGGAT 662
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 GCAAGGCACGCCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTTCCCGCTGGGAT 1028

Qy     663 CCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAGGAGGGCTTGACCATAAAGAA 722
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1029 CCTGGTCCCCACCCAGTTGTTGGTGCCATCATCGGAAGGAGGGCTTGACCATAAAGAA 1088

Qy     723 CATCACTAAGCAGAGACCCAGTCCCGGGTAGATATCCAGAGTATCCAGAGAACTCTGGAGCTGC 782
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1089 CATCACTAAGCAGAGACCCAGTCCCGGGTAGATATCCAGAGTATCCAGAGAACTCTGGAGCTGC 1148

Qy     783 AGAGAAGCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGAT 842
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1149 AGAGAAGCCTGTCACCATCATGCACCCCAGAGGGGACTTCTGAAGCATGCCGCATGAT 1208

Qy     843 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 902
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1209 TCTTGAAATCATGCAGAAAGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy     903 AATCTTGGCACACAATGGCTTGGTTGGAGACTGATTGGAAGACTGATTGGAAGAGGCAGAAATTTGAA 962
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1269 AATCTTGGCACACAATGGCTTGGTTGGAGACTGATTGGAAGACTGATTGGAAGAGGCAGAAATTTGAA 1328

Qy     963 GAAAATTGAACATGAAACAACCATCACTGTGAGGCACAGTTGAGGCTTTGAAAATGATATGCTGCTGTTAACCA 1022
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db    1329 GAAAATTGAACATGAAACAACCATCACTGTGAGGCACAGTTGAGGCTTTGAAAATGATATGCTGCTGTTAAC-- 1388

Qy    1023 ATACAACCCGAAAGAACCATCACTGTGAGGCACAGTTGAGGCTTGAGGGCCACAGTTGCCAGTGCTGA 1082
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1389 ATACAACCCGAAAGAACCATCACTGTGAGGCACAGTTGAGGCTTGAGGGCCACAGTTGCCAGTGCTGA 1448

Qy    1083 GATAGAGATTATGAAGAAGCTGCGTGAGGCTTTGAAAATGATATGCTGCTGTTAACCA 1142
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    1449 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGCTGTTAAC-- 1506

Qy    1143 ACAAGCCAATCTGATCCCAGGGTTGAACCTTCAGCGCACTTGGCATCTTTTCAACAGGACT 1202
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1507 --------------------------------------------------- 1506

Qy    1203 GTCCGTGCTATCTCCACCACCAGCAGCAGGGCCCCGCGGAGCTCCCCGCTGCCCCTACCACCC 1262
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1507 --------------------------------------------------- 1506

Qy    1263 CTTCACTACCACTCCGGATACTTCTCCAGCAGGAGAGGGGCACCATCAAACAGCTGGGAGATTCGGCCGTT 1322
           |||||||||||||||||||||||||||||||||
Db    1507 ------CCCAGTACTTCTCCAGCAGGAGAGTCGGGGCACCATCAAACAGCTGGGAGATTCGGCCGTT 1559

Qy    1323 CCCGCATCATCACTCTATCCGAGCAGGAGGATTGTGAATCTCTTCATCCCAACCCAGGC 1382
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCCGCATCATCACTCTTATCCGAGCAGGAGGATTGTGAATCTCTTCATCCCAACCCAGGC 1619

Qy    1383 TGTGGGCGCCATCATCGGGAAGAGAGGGGGCACACATCAAACAGCTGGGAGATTCGCCGG 1442
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 TGTGGGCGCCATCATCGGGAAGAGAGGGGGCACACATCAAACAGCTGGGAGATTCGCCGG 1679

Qy    1443 AGCTCTATCAAGATTGCCCCTGCGAAGGCCCAGACGTCAGCGGAAAGGATGGTCATCAT 1502
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 AGCCTCTATCAAGATTGCCCCTGCGAAGGCCCAGACGTCAGCGGAAAGGATGGTCATCAT 1739

Qy    1503 CACCGGGCCACCGGGAGCCCAGTTCAAGGCCCAGGAGGACGACGGATCTTTGGAAACTGAAAGA 1562
```

```
Qy       3 CAGCCGGAGGAGGCGAGGAGGCCGAGGAGCCCGGGTACCGGCCCGGGAGCCGGGGAGCCCTCCGGGGAA 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 CAACGGAGGAGGCGAGGAGGCGAGGAGCCCGGGTACCGGCGGGGAGCCGGGGAGCCCTCCGGGGAA 428

Qy      63 GAGACCGGATGATGAACAAGCTTTACATCGGGAACAAGCTGAGCCCGTGCACCGCGCCGACGA 122
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 GAGACCGGATGATGAACAAGCTTTACATCGGGAACAAGCTGAGCCCGCGGACACCGCGCCGACGA 488

Qy     123 CCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTGGCCCCGGCGGGACAGGTCCTGCTGAAGTC 182
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 CCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTGGCCCCGGCGGGACAGGTCCTGCTGAAGTC 548

Qy     183 CGGCTACGCCTTCGTGACTACCCCGACCACCAGAACTGGGCCATCGCCGCCATCGAGACCCT 242
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 CGGCTACGCCTTCGTGACTACCCCGACCACCAGAACTGGGCCATCGGCCCATCGAGACCCT 608

Qy     243 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA 302
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAAA 668

Qy     303 GCTAAGGAGCAGGAGAAATTCAGATTCGAAAACATCCTCCTCACCTTGCAGTTGGGAGGTGTT 362
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 GCTAAGGAGCAGGAGAAAATTCAGATTCGAAAACATCCTCCTCACCTTGCAGTTGGGAGGTGTT 728

Qy     363 GGATGCAGCTTTTGGCTCAATATGGGACCAGTGGAACAAGTGGAATGTGGAAACTGAAAGA 422
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     729 GGATGCAGCTTTTGGCTCAATATGGGACCAGTGGAACAAGTGGAATGTGGAAACTGAAAGA 788
```

```
Db  1740  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  1799
          CACCGGCCACCGGAAGCCAGTTCAAGGCCGATCGTTCAAGGACCGGATCTTTGGGAAACTGAAAGA

Qy  1563  GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCGCCATATCAGAGTGCCCTC           1622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1800  GGAAAACTTCTTTAACCCCAAAGAAGAAGTGAAGCTGAAGCGCCATATCAGAGTGCCCCC          1859

Qy  1623  TTCCACAGCTGCCCGGTGATTGGCAAGACCGTGAAGACGACTGCAGAACTT                    1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1860  TTCCACAGCTGCCCGGTGATTGGCAAGACCGTGAAGACGACTGCAGAACTT                   1919

Qy  1683  AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGATGAAAATGAGGAAGTGAT                  1742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1920  AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAGATGAAAATGAGGAAGTGAT                 1979

Qy  1743  CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGGAAAT         1802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1980  CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGACTGCACAGCCCAAGATCAGGGGAAAT        2039

Qy  1803  TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG          1862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2040  TGTACAACAGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG         2099

Qy  1863  CAAGTGAGGCTCCCACAGCCACCAGCCAAACAACGATGAATGTAGCCCTTCCAACACCT          1922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2100  CAAGTGAGGCTCCCACAGCCACCAGCCAAACAACGATGAATGTAGCCCTTCCAACACCT         2159

Qy  1923  GACAGAATGAGACCAAACGACCAGCCAGCAAACCAAAGACCATCTGAGA                     1982
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2160  GACAGAATGAGACCAAACGACCAGCCAGCAAACCAAAGACCATCTGAGA                    2219

Qy  1983  ATGAGAAGTCTGCGGAGGCGGCGGAAGTCAGGGCTCGCCAGGGACTCGCCGAGGGCGCG          2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2220  ATGAGAAGTCTGCGGAGGCGGCGGAAGTCAGGGCTCGCCAGGGACTCGCCGAGGGCGCG         2279

Qy  2043  AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCCAGGAACCACCACCATCCACTCCGAGATC         2102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2280  AGGAGGGGCGGGGAAGGTCAGCCAGGTTGCCCAGGAACCACCACCATCCACTCCGAGATC        2339

Qy  2103  CCAGGGCTTCTGCAGGCTTCAGCTCAGCCTTCAGCCATCATCCACTTCACCTGGATCTCTCCTGAACT  2162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2340  CCAGGGCTTCTGCAGGCTTCAGCTCAGCCATCATCCACTTCACCTGGATCTCTCCTGAACT       2399

Qy  2163  CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA         2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2400  CCCACGACGCTATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA        2459

Qy  2223  AATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTACATATTAGAAAGG             2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2460  AATGCACACCCTTTTCTGTGGCAAATCGTCTGTACATGTGTACATATTAGAAAGG            2519

Qy  2283  GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCGACGGTAATATATTT          2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2520  GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGTGCCTGCGACGGTAATATATTT         2579

Db  2343  TAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTTTT                     2402
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2580  TAGAAATAATATATCAAATAACTCAATTTTAATCAATTATTAATTTTTT                    2639

Db  2403  TTTCTTTTAAGAGAAGACAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT          2462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2640  TTTCTTTTAAGAGAAGACAGCCTTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT         2699

Db  2463  CTCACCGTGTAGAGAAGAGCTTTGAGCCACCCGACACAAATTCACCCGAGGGGAAATCT          2522
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2700  CTCACCGTGTAGAGAAGAGCTTTGAGCCACCCGACACAAATTCACCCGAGGGGAAATCT         2759

Db  2523  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGGATACC              2582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2760  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACCTGTATGTCAACAGAGGATACC           2819

Db  2583  GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCGCGTGTCTCAGTCTCATACACCCATTTC   2642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2820  GTCTCCTTGAAGAGGAAACTCTGTCACTCCTCCTCATGCGCGTGTCTCAGTCTCATACACCCATTTC  2879

Qy  2643  TCTTTGCTTCACAGGTTTAAACTGGTTTTTGCATACTGCTATATAATTCTCTGTCTCT           2702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2880  TCTTTGCTTCACAGGTTTAAACTGGTTTTTTGCATACTGCTATATAATTCTCTGTCTCT        2939

Qy  2703  CTCTGTTTATCTCTCCATTCTTCTCCATCTCCATTCTTTGAATT                         2762
          ||||||||||||||||||||||||||||||||||||||||
Db  2940  CTCGTTTATCTCTCCATTCTTCTCCATCTCCATTCTTTGAATT                         2999

Qy  2763  TCCTCATCCCTCCATCCAATCCCGTATCTCACGCACCCCCCCCCCCCAGGCAAAGCAGT          2822
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  3000  TCCTCATCCCTCCATCTCAATCCCGTATCTACGGCA -CCCCCCCCCCCCAGGCAAAGCAGT       3058

Qy  2823  GCTCTGAGTATCACATCCACACAAGAGGACAAAAGCGAAACACAAACCAGCCTCAACT          2882
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3059  GCTCTGAGTATCACATCCACACAAGAGGACAAAAGCGAAACACAAACCAGCCTCAACT         3118

Qy  2883  TACACTTCGTTACTCAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG                   2942
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3119  TACACTTGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTGGAAGAG          3178

Qy  2943  GAAAACAGGAACCCACCAAACAACCATCACCAAACAAAGAAAAATTCCACAATGAA             3002
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3179  GAAAACAGGAACCCACCAAACAACCAACCATCACCAAACAAAGAAAAAATTCCACAATGAA      3238

Qy  3003  AGAATGATTTTGTCTTTTGCATTTTGGTGTATAAGGCCATCAATATTCAGCAAAATGAT         3062
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3239  AGAATGTATTTTGTCTTTTGCATTTTGGTGTGTTATAAGCCATCAATATTCAGCAAAATGAT    3298

Qy  3063  TCCTTTCTTT-AAAAAAAAAAATGTGAGGAAGTAGAAAATTACCAAGGTTGTTGGCCC          3121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3299  TCCTTTCTTTAAAAAAAAAAAAATGTGAGGAAGTAGAAAATTACCAAGGTGTTTGGCCC       3358

Qy  3122  AGGGCGTAAATTCACAGATTTTTTTAACGAGCAAAACACACAGAGAAGAGCTACCTTCAGG       3181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3359  AGGGCGTAAAATTCACAGATTTTTTTTAACGAGCAAAACACACAGAGAAGAGCTACCTTCAGG    3418

Qy  3182  TGTTTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGT          3241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3419  TGTTTTTACCTCAGCACCTTGCTCTTGTTTCCCTTAGAGATTTGTAAAGCTGATAGT         3478

Qy  3242  TGGAGCATTTTTTATTTTTAATAAAAATGAGTTGGAAAAAATAAGATATCAACTG             3301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3479  TGGAGCATTTTTTTATTTTTAATAAAAATGAGTTGGAAAAAAATAAGATATCAACTG          3538

Qy  3302  CCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTA        3361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3539  CCAGCCTGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTA       3598

Qy  3362  GCCAAGAACCNATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT         3412
          ||||||||||||||||||||||||||||||||||||||||||||
Db  3599  GCCAAGAACCTATATGGCCTTCTTTTTGGACAAACCTTGAAAATGTTTATTT        3649
```

```
;   APPLICANT: Rosemarie SCHMANDT
;   APPLICANT: Xumei ZHAO
;   APPLICANT: Karen GLATT
;   TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;                       Assessment, Prevention, and Therapy of Ovarian Cancer
;   FILE REFERENCE: MRI-030
;   CURRENT APPLICATION NUMBER: US/10/097,340
;   CURRENT FILING DATE: 2002-03-14
;   PRIOR APPLICATION NUMBER: 60/276,025
;   PRIOR FILING DATE: 2001-03-14
;   PRIOR APPLICATION NUMBER: 60/325,149
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: 60/276,026
;   PRIOR FILING DATE: 2001-03-14
;   PRIOR APPLICATION NUMBER: 60/324,967
;   PRIOR FILING DATE: 2001/09/26
;   PRIOR APPLICATION NUMBER: 60/311,732
;   PRIOR FILING DATE: 2001-08-10
;   PRIOR APPLICATION NUMBER: 60/325,102
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: 60/323,580
;   PRIOR FILING DATE: 2001-09-19
;   NUMBER OF SEQ ID NOS: 363
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 146
;   LENGTH: 3667
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-097-340-146

Query Match             91.3%;  Score 3114.8;  DB 15;  Length 3667;
Best Local Similarity   96.1%;  Pred. No. 0;
Matches 3277;  Conservative 0;  Mismatches 0;  Indels 131;  Gaps 3;
```

```
Qy    3 CAGCGGAGGAGCGAGGAGCGCGGTACCGGCCCGGGAGCCCGGGAGCTGCTTCGGGGAA  62
          ||||||||||| |||||||||  ||||||  |||||||||||||||||||||||||||
Db  369 CAACGGAGGAGGCGAGGAGCCCGGGTACCGGGCCCGGGAGCCCGGGAGCTGCTTCGGGGAA 428

Qy   63 GAGACCGATGATGAACAAGCTTTACATCGGGAAACTCGGGAAACCTGAGCCCGCCGACGA 122
          ||||||||||||||||||| |||||| |||| |||| |||||||||||||||||||||
Db  429 GAGACCGATGATGAACAAGCAGTTTACATCGGGAAACCTGAGCCCGCCGACGA 488

Qy  123 CCTCCCGCCAGTCTTTGGGACAGGAAGCTGCCCCGGCGGCCATCGGGCACAGGTCCTGCTGAAGTC 182
          |||||||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db  489 CCTCCCGCCAGTCTTTGGGACAGGAAGCTGCCCCGGCGGCCATCGGGCACAGGTCCTGCTGAAGTC 548

Qy  183 CGGCTACGCCTTCGTGACTGACTACCCCGACCAGGCCATCGGCCATCGGCCATCTTTTCAACAGGACT 242
          |||||||||||||||||||||||| |||||||||||||||||  |||  |||||||||||||||
Db  549 CGGCTACGCCTTCGTGACTGACTACCCCGACCAGGCCATCGGCCATCGGCCATCTTTTCAACAGGACT 608

Qy  243 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAA 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  609 CTCGGGTAAAGTGGAATTGCATGGGAAAATCATGGAAGTTGATTACTCAGTCTCTAAAA 668

Qy  303 GCTAAGGAGCAGGAAATTCAGATTCGAAAATTCGAAACATCCCTCCTCACCTGCAGTGGGAGTGTT 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  669 GCTAAGGAGCAGGAAATTCAGATTCGAAAATTCGAAACATCCCTCCTCACCTGCAGTGGGAGTGTT 728

Qy  363 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACAC 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  729 GGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGAACAAGTCAACACAGACAC 788

Qy  423 AGAAACCGCCGGTTGTCACGTCACGTCACATATGCACAGAGAAGAGCAAAAATAGCCATGGA 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  789 AGAAACCGCCGGTTGTCACGTCACATATGCACAGAGAAGAGCAAAAATAGCCATGGA 848

Qy  483 GAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAGTTCCTTACATCCCGATGA 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  849 GAAGCTAAGCGGGCATCAGTTTGAGAACTACTCCTTCAAGAGTTCCTTACATCCCGATGA 908

Qy  543 AGAGGTGAGCTCCCCTTCGCCCCCTCAGCGCAGCCCAGCGCGTGGGGACCACTCTTCCCGGGA 602
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  909 AGAGGTGAGCTCCCCTTCGCCCCCTCAGCGCAGCCCAGCGCGTGGGGACCACTCTTCCCGGGA 968
```

```
Qy  603 GCAAGGCCACGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCGGTGCGGAT 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  969 GCAAGGCCACGCGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCCGGTGCGGAT 1028

Qy  663 CCTGTCTCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1029 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAA 1088

Qy  723 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAGAAAGAGAACTCTGGAGCTGC 782
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 1089 CATCACTAAGCAGACCCAGTCCCGGTAGATATCCATAGAAAAGAGAACTCTGGAGCTGC 1148

Qy  783 AGAGAAGCCTGTCACCATGCACCCCAGAGGGGGACTCTGAAGCATGCCGCCATGAT 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1149 AGAGAAGCCTGTCACCATGCACCCCAGAGGGGGACTCTGAAGCATGCCGCCATGAT 1208

Qy  843 TCTTGAAATCATGCAGAAGAGGCAGATGAGGACCAAACTAGCCGAAGAGATTCCTCTGAA 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1209 TCTTGAAATCATGCAGAAGAGGCAGATGAGGACCAAACTAGCCGAAGAGATTCCTCTGAA 1268

Qy  903 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGAGGCAGAAATTTGAA 962
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1269 AATCTTGGCACACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAGGAGGCAGAAATTTGAA 1328

Qy  963 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCAT 1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1329 GAAAATTGAACATGAAACAGGGACCAAGATAACAATCTCATCTTTGCAGATTTGAGCAT 1388

Qy 1023 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCTGTGCCAGTGCTGA 1082
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1389 ATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTTGAGGCTGTGCCAGTGCTGA 1448

Qy 1083 GATAGAGATTATGAGAAGCTGCGTAGCCTTTGAAAATGATATGATATGCTGGTGTTAACCA 1142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||  --
Db 1449 GATAGAGATTATGAGAAGCTGCGTGAGCCTTTGAAAATGATATGATATGCTGGCTGTTAAC-- 1506

Qy 1143 ACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGACT 1202
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1507 ACAAGCCAATCGATCCCAGGGTTGAACCTCAGCGGCACTTGGCATCTTTTCAACAGGACT 1506

Qy 1203 GTCCGGTGCTATCTCCACCAGCAGGTCCTGCTGAAGTC 1262
          ----------------------------------
Db 1507 ---------------------------------- 1506

Qy 1263 CTTCACTACCCACTCCGATACTTCCCAGCCTGTACCCCATCACCAGTTTGGCCGCCGTT 1322
          ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db 1507 ------ACCCACTCCGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCGCCGTT 1559

Qy 1323 CCCGGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCAACCCAGGC 1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1560 CCCGGCATCATCACTCTTATCCAGAGCAGGAGATTGTGAATCTTCTTCAACCCAGGC 1619

Qy 1383 TGTGGGCGCCATCATCGGGAAGAGGGGGCACACATCAAACAGTCGGCGAATTCGCCGG 1442
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1620 TGTGGGCGCCATCATCGGGAAGAGGGGGCACACATCAAACAGTCGGCGAATTCGCCGG 1679

Qy 1443 AGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT 1502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1680 AGCCTCTATCAAGATTGCCCCTGCGGAAGGCCCAGACGTCAGCGAAAGGATGGTCATCAT 1739

Qy 1503 CACCGGCCACCGAAGCCCAGGAGCCCAGGGACGGATCTTTGGGAACTGAAAGA 1562
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1740 CACCGGCCACCGAAGCCCAGGAGCCCAGGGACGGATCTTTGGGAACTGAAAGA 1799

Qy 1563 GGAAAACTTCTTTAACCCCAAAGATTCCTTACATCCCGATGA 1622
          |||||||||||||||||||||||||||||||||||||||||
Db 1800 GGAAAACTTCTTTAACCCCAAAGATTCCTTACATCCCGATGA 1859

Qy 1623 TTCCACAGCTGGCCGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTT 1682
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1860 TTCCACAGCTGGCCGGTGATTGGCAAGGTGGCAAGACCGTGAACGAACTGCAGAACTT 1919
```

```
Qy   1683  AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT  1742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920  AACCAGTGCAGAAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAGTGAT  1979

Qy   1743  CGTCAGAATTATCGGGACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT   1802
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980  CGTCAGAATTATCGGGACACTTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAAT  2039

Qy   1803  TGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  1862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040  TGTACAACAGGTGAAGCAGCAGGAGCAGAAATACCCTCAGGGAGTCGCCTCACAGCGCAG  2099

Qy   1863  CAAGTGAGGCTCCCACAGGGACCAGCAACAACGGATGAATGTAGCCCTTCCAACACCT    1922
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100  CAAGTGAGGCTCCCACAGGGACCAGCAACAACGGATGAATGTAGCCCTTCCAACACCT    2159

Qy   1923  GACAGAGATGAGACCAGACCAGCCAGCAGCAGACCAAACCAGAGCCATCTGAGGA        1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160  GACAGAGATGAGACCAGACCAGCCAGCAGCAGACCAAACCAGAGCCATCTGAGGA        2219

Qy   1983  ATGAGAAGTCTGCGGAGGCGGGCAAGTCAGTTGCCAGGACTCTGCCCAGGGCCG        2042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220  ATGAGAAGTCTGCGGAGGCGGGCAAGTCAGTTGCCAGGACTCTGCCCAGGGCCG        2279

Qy   2043  AGGAGGGGCGGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCCCC   2102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280  AGGAGGGGCGGGGGAAGTCAGCCAGTTTGCCAGAACCACCGAGCCCCGCCTCCCGCCCC   2339

Qy   2103  CCAGGGCTTCGCAGCCATCCACTCAGCGCCATCCACTCCCTGAACT                2162
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   2340  CCAGGGCTTCGCAGCCATCCACTCAGCGCCATCCACTCCCTGAACT                2399

Qy   2163  CCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA   2222
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400  CCCACGACGCTATCCCTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA   2459

Qy   2223  AATGCACACCCTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGG     2282
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2460  AATGCACACCCTTTCTGTGGCAAATCGTCTCTGTACATGTGTACATATTAGAAAGG     2519

Qy   2283  GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATATTT  2342
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2520  GAAGATGTTAAGATATGTGGCCTGTGGGTTACACAGGGTGCCTGCAGCGTAATATATTT  2579

Qy   2343  TAGAAATAATAATATCAAATAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT   2402
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2580  TAGAAATAATAATATCAAATAACTGTTTTAAACTCAACTAACTCCAATTTTAATCAATTATTAATTTT   2639

Qy   2403  TTTCTTTTAAAGAGAAAGCAGAGGCTTTCTAGACTTAAAGAATAAAAGTCTTTGGGAGT  2462
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2640  TTTCTTTTAAAGAGAAAGCAGAGGCTTTCTAGACTTAAAGAATAAAAGTCTTTGGGAGT  2699

Qy   2463  CTCACGGTGTAGAGAGGAGCTTGAGGCACCCGCACAAAATTCACCCAGAGGGAAATCT   2522
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2700  CTCACGGTGTAGAGAGGAGCTTGAGGCACCCGCACAAAATTCACCCAGAGGGAAATCT   2759

Qy   2523  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACGTGTATGTCAACAGAAGGGATACC   2582
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2760  CGTCGGAAGGACACTCACGGCAGTTCTGGATCACGTGTATGTCAACAGAAGGGATACC   2819

Qy   2583  GTCTCCTTGAAGGAAACTCTGTCACTCCTCAGCTCATACACCCATTTC             2642
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   2820  GTCTCCTTGAAGGAAACTCTGTCACTCCTCAGCTCATACACCCATTTC             2879

Qy   2643  TCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGGTTTTTGCATACTCTCTGTCTCT  2702
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2880  TCTTTGCTTCACAGGTTTTAAACTGGTTTTTGCATACTGGTTTTTGCATACTCTCTGTCTCT  2939

Qy   2703  CTCTGTTTATCTCTCCCCTCTTCCCCCTCCCCCTCCCCATCTTCCATTCTTTTGAATT    2762
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2940  CTCTGTTTATCTCTCCCCTCTTCCCCCTCCCCCTCCCCATCTTCCATTCTTTTGAATT    2999

Qy   2763  TCCTCCATCCCTCCATCTCAATCCCGTATCTACGGACCCCCCCCCCCGGGCCAAAGCAGT  2822
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3000  TCCTCATCCTCCATCTCAATCTCCGTATCTCACGGCA-CCCCCCCCCCCAGGGCAAAGCCTCAACT  3058

Qy   2823  GCTCTGAGTATCACATCACACAAAAGGAACAAAAAGCGAAACACACAACACAGCCTCAACT  2882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3059  GCTCTGAGTATCACATCACACAAAAGGAACAAAAAGCGAAACACACAACACAGCCTCAACT  3118

Qy   2883  TACACTTGGTTACTCAAAAGAACAAGAGAGTCAATGGTACTTGTCTCCTAGCGTTTTGGAAGAG  2942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3119  TACACTTGGTTACTCAAAAGAACAAGAGAGTCAATGGTACTTGTCTCCTAGCGTTTTGGAAGAG  3178

Qy   2943  GAAAACAGGAACCCACCAAACCAACCAATCAACCAATCACCCAAACAAAGAAAAATTCCAACAATGAA  3002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3179  GAAAACAGGAACCCACCAAACCAACCAATCAACCAATCACCCAAACAAAGAAAAATTCCAACAATGAA  3238

Qy   3003  AGAATGTATTTTGTCTTTGCATTTTGTGTTGGTATAAGCCATCAATATTCAGCAAAATGAT  3062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3239  AGAATGTATTTTGTCTTTGCATTTTGTGTTGGTATAAGCCATCAATATTCAGCAAAATGAT  3298

Qy   3063  TCCTTTCTTT-AAAAAAAAAAATGTGAGGAAGAAATTACCAAGGTTGTTGGCCC        3121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299  TCCTTTCTTTAAAAAAAAAAAATGTGAGGAAGAAATTACCAAGGTTGTTGGCCC        3358

Qy   3122  AGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGG   3181
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359  AGGGCGTTAAATTCACAGATTTTTTAACGAGAAAAACACACAGAAGAAGCTACCTCAGG   3418

Qy   3182  TGTTTTTACCTCAGCACCTTGCTCCTTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGT  3241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419  TGTTTTTACCTCAGCACCTTGATCTTGTTTCCCTTAGAGATTTTGTAAAGTGATAGT   3478

Qy   3242  TGGAGCATTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAATATCAACTG  3301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3479  TGGAGCATTTTTTTATTTTTTTAATAAAAATGAGTTGGAAAAAAATAAGAATATCAACTG  3538

Qy   3302  CCAGCCTCGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTA  3361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3539  CCAGCCTCGGAGAAGGTGACAGTCCAAGTGTGCAACAGCTGTTCTGAATTGTCTTCCGCTA  3598

Qy   3362  GCCAAGAACCNATATGGCCTTCTTTTTGACAAACCTTGAAAATGTTTATTT          3412
           ||||||||| |||||||||||||||||||||||||| ||||||||||||||
Db   3599  GCCAAGAACCTATATGGCCTTCTTTTTGACAAACCTTGAAAATGTTTATTT          3649
```

```
Query Match             91.3%;  Score 3114.8;  DB 17;  Length 3667;
Best Local Similarity   96.1%;  Pred. No. 0;
Matches  3277;  Conservative    0;  Mismatches    3;  Indels  131;  Gaps    3;

Qy      3  CAGCGGAGGAGGCGAGGAGGCGGAGGCGCCGGGTACCCGGGGGAGCCTCCGGGGAA     62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369  CAACGGAGGAGGCGGAGGAGGCGGAGGCGCCGGGTACCCGGGGGAGCCTCCGGGGAA    428
```

```
Qy    63 GAGACGCGGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGTCACCGGCGACGA 122
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   429 GAGACGCGATGATGAACAAGCTTTACATCGGGAACCTGAGCCCCGCCGTCACCGGCGACGA 488

Qy   123 CCTCCGCGAGCTCTTGGGACACGGAGAGCTGCCCCTGCGGGGACAGGTCCTGCTGAAGTC 182
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CCTCCGCGAGCTCTTGGGACACGGAGAGCTGCCCCTGCGGGGACAGGTCCTGCTGAAGTC 548

Qy   183 CGGCTAGCCTTCGTGGACTACCCCGACCAGAACTCGGGCCATCCGGGGCCATCCGAGACCCT 242
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   549 CGGCTAGCGCTTGTGGACTACCCCGACCAGAACTCGGGCCATCCGGGCCATCCGAGACCCT 608

Qy   243 CTCGGGTAAAGTGGCTCAATATGGACAGTGGACGAATGCATGTGGAAAATCATGGAGTTGATTACTCAGTCTCTAAAAA 302
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   609 CTCGGGTAAAGTGGCTCAATATGGACAGTGGACGAATGCATGTGGAAAATCATGGAGTTGATTACTCAGTCTCTAAAAA 668

Qy   303 GCTTAAGGAGCAGGAAAATTCAGATTCGAAAACATCCCTCCTACCTGCAGTGGGAGGTGTT 362
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669 GCTTAAGGAGCAGGAAAATTCAGATTCGAAAACATCCCTCCTACCTGCAGTGGGAGGTGTT 728

Qy   363 GGATGGACTTTTGGCTCAATATGGACAGTGAGAATGTGGACAAGTCAACACAGACAC 422
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   729 GGATGGACTTTTGGCTCAATATGGACAGTGAGAATGTGGACAAGTCAACACAGACAC 788

Qy   423 AGAAAACCGCCGTTGTCAAGTCACATATGCAACAAGAGAAGAGCAAAAAATAGCCATGGA 482
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   789 AGAAAACCGCCGTTGTCAAGTCACATATGCAACAAGAGAAGAGCAAAAAATAGCCATGGA 848

Qy   483 GAAGCTAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA 542
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   849 GAAGCTAAGCGGGGCATCAGTTTGAGAACTACTCCTTCAAGATTTCCTACATCCCGGATGA 908

Qy   543 AGAGTGAGCTCCCCCTCGCCCCCTCAGCGAGCCCAGGCCACCTCTTCCCGGGA 602
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909 AGAGGTGAGCTCCCCCTCGCCCCCTCAGCGAGCCCAGGCCACCTCTTCCCGGGA 968

Qy   603 GCAAGGCCACGCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCCGTCGCGGAT 662
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   969 GCAAGGCCACGCCCCTGGGGCACTTCTCAGCGCCAGACAGATTGATTTCCCGTCGCGGAT 1028

Qy   663 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGCATCATCGGAAAGAGAGGCTTGACCATAAAGAA 722
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1029 CCTGGTCCCCACCCAGTTTGTTGGTGCCATCATCGGGCATCATCGGAAAGAGAGGCTTGACCATAAAGAA 1088

Qy   723 CATCACTAAGCAGCAGACCCCAGTCCCGGGTAGATATCCATAGAGAAAAGAGAACTCATAGAGAAAAGAGACTCTGGAGCTGC 782
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1089 CATCACTAAGCAGCAGACCCCAGTCCCGGGTAGATATCCATAGAGAAAAGAGAACTCTGGAGCTGC 1148

Qy   783 AGAGAAGCCTGTCACCATCCACCATCCACGAGGGGACTTCTGAAGCATGCGCATGAT 842
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1149 AGAGAAGCCTGTCACCATCCACCATCCACGAGGGGACTTCTGAAGCATGCGCATGAT 1208

Qy   843 TCTTGAAATCATGCACACAATGGCTTGGTTGGAAGACTACGCCGAAGATTCCTCTGAA 902
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1209 TCTTGAAATCATGCACACAATGGCTTGGTTGGAAGACTACGCCGAAGATTCCTCTGAA 1268

Qy   903 AATCTTGCCACACAATGAAACATGAAAACATGAAAACAATCTCATCTTTGCAGAAATTTGAA 962
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1269 AATCTTGCCACACAATGAAACATGAAAACAATCTCATCTTTGCAGGATTTGAGCAT 1328

Qy   963 GAAAATTGAACAATGAAAACATGAAACAAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1022
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1329 GAAAATTGAACAATGAAACAAGGACCAAGATAACAATCTCATCTTTGCAGGATTTGAGCAT 1388

Qy  1023 ATACAACCCGGAAAGAGACCATCACTGTGGAAGGGCCAGTGGCAGTCTGTGCCAGTGCTGA 1082
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1389 ATACAACCCGGAAAGAGACCATCACTGTGGAAGGGCCAGTGGCAGTCTGTGCCAGTGCTGA 1448

Qy  1083 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAACCA 1142
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1449 GATAGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATATGCTGGCTGTTAAC-- 1506
```

```
Qy  1143 ACAAGCCAATCTGATCCAGGTTGAACCTCAGCGCGCACTTGGCATCCTTTCAACAGGACT 1202
         ||||||||||
Db  1507 ---------------------------------------------------------- 1506

Qy  1203 GTCCGTTGCTATCTCCACCAGCAGGCCCCGCGGAGCTCCCCCCGTGCCCCTACCACCC 1262
         |||||||||||||||
Db  1507 ---------------------------------------------------------- 1506

Qy  1263 CTTCACTACCCACTCCGATACTTCTCCAGCCTGTACCCCATCACCAGTTTGGCCCGTT 1322
         ||||||||||||||||||        |||||||||||||||||||||||||||||||
Db  1507 ------ACCCACTCCGGTATCTTCT--CAGCCTGTACCCCATCACCAGTTTGGCCCGTT 1559

Qy  1323 CCCGGATCATCACTCTTATCCAGAGCAGGAGATTGTGAAATCTTCATCCCAACCCAGGC 1382
         ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1560 CCCGGCATCATCATCTTATCCAGAGCAGGAGATTGTGAATCTTCATCCCAACCCAGGC 1619

Qy  1383 TGTGGGCGCCATCATCGGGAAGAGGGGGCACCATCAACATCAAACAGCTGGCGAGATTCGCCGG 1442
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1620 TGTGGGCGCCATCATCGGGAAGAGGGGGCACCATCAACATCAAACAGCTGGCGAGATTCGCCGG 1679

Qy  1443 AGCCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGAACGTCAGCGAAAGGATGGTCATCAT 1502
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1680 AGCCCTCTATCAAGATTGCCCCTGCCGGAAGGCCCAGAACGTCAGCGAAAGGATGGTCATCAT 1739

Qy  1503 CACCCGGCCACCGAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGA 1562
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1740 CACCCGGCCACCGGGAGCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTGAAAGA 1799

Qy  1563 GGAAAACTTCTTTAACCCCAAAGAAGAGATTTCCTACATCCCGGATGA 1622
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1800 GGAAAACTTCTTTAACCCCAAAGAAGAGAAGTGAAGCTGGAAGACCGTGAACCTGCAGAACTT 1859

Qy  1623 TTCCACAGCTCGCCGGGTGATTGGCAAGACCGTGAACCTGCAGAACTT 1682
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1860 TTCCACAGCTCGGCCGGGTGATTGGCAAGACCGTGAACCTGCAGAACTT 1919

Qy  1683 AACCAGTGCAGAAGTCATCGTGCCCGTGACCAAAACGCCAGATGAAAATAGAGGAAGTGAT 1742
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1920 AACCAGTGCAGAAGTCATCGTGCCCGTGACCAAAACGCCAGATGAAAATAGAGGAAGTGAT 1979

Qy  1743 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGCTGCACAGCGGCAAGATCAGGGAAAT 1802
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1980 CGTCAGAATTATCGGGCACTTCTTTGCTAGCCAGCTGCACAGCGGCAAGATCAGGGAAAT 2039

Qy  1803 TGTACAACAGTGAAGCAGCAGAGAATACCCTCAGGGAGTCGCCTCACACGCGCAG 1862
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2040 TGTACAACAGTGAAGCAGCAGAGAATACCCTCAGGGAGTCGCCTCACACGCGCAG 2099

Qy  1863 CAAGTGAGGCTCCCACAGGACCAGCAACAACAACGGATGAATGAATGTAGCACACCT 1922
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2100 CAAGTGAGGCTCCCACAGGACCAGCAACAACAACGGATGAATGAATGTAGCCACACCT 2159

Qy  1923 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGA 1982
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2160 GACAGAATGAGACCAAACGCAGCCAGCCAGATCGGGAGCAAACCAAAGACCATCTGAGGA 2219

Qy  1983 ATGAGAAGTCTGCGCAGGGGCCGGCCGGGGAGGCCGGCCTGAGGCCCTGAGAGCCCCGCCG 2042
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2220 ATGAGAAGTCTGCGCAGGGGCCGGCCGGGGAGGCCGGCCTGAGGCCCTGAGAGCCCCGCCG 2279

Qy  2043 AGGAGGGGCGGGGAAGGTCAGCGAAGGTCAGCGAGGACTCTGCGAGGGCTCGGGGGCCG 2102
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2280 AGGAGGGGCGGGGAAGGTCAGCGAAGGTCAGCGAGGACTCTGCGAGGGCTCGGGGGCCG 2339

Qy  2103 CCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCTCACCTCGGATCTCTCGGAACT 2162
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2340 CCAGGGCTTCTGCAGGCTTCAGCCATCCACTTCACCTCACCTCGGATCTCTCCTGAACT 2399

Qy  2163 CCCACGACGTATCCCTTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2222
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2400 CCCACGACGCATCCCTTTTAGTTGAACTAACATAGGTGAACGTGTTCAAAGCCAAGCAA 2459

Qy  2223 AATGCACACCCTTTTCTGTGGCAAATCGTCTCTGTACATGTGTGTACATATTAGAAAGG 2282
```

```
Db  2460  AATGCACACCCTTTTCTGTGCAAATCGTCTCTGTACATGTGCTGCTCAGCGGTAATATTAGAAAGG  2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3539  CCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTGTCTGAATTGTCTTCCGCTA  3598

Qy  2283  GAAGATGTTAAGCATGTAGATATGTGGGCTGCTGGTTACACAGGGTGCTCGAGCGGTAATATATT  2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3362  GCCAAGAACCNATATGGCCTTCTTTTTGACAAACCTTGAAAAATGTTTATTT  3412

Db  2520  GAAGATGTTAAGCATGTAGATATGTGGGCTGCTGGTTACACAGGGTGCTCGAGCGGTAATATATT  2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3599  GCCAAGAACCTATATGGCCTTCTTTTTTGACAAACCTTGAAAATGTTTATTT  3649
```

RESULT 7
US-09-764-864-332
; Sequence 332, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
; LENGTH: 2290
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-332

Query Match   58.3%; Score 1989.8; DB 9; Length 2290;
Best Local Similarity   99.4%; Pred. No. 0;
Matches 2028; Conservative 0; Mismatches 0; Indels 8; Gaps 3;

```
Qy  1376  CCCAGGCTGTGGGCCGCCATCATCGGGAGGAAGGGGCACACATCAAACAGCTGGCGACGAGAT  1435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25  CCCACGGCTCCGGCCCATCATCGGGAGGAAGGGGCACACATCAAACAGCTGGCGACGAGAT    84

Qy  1436  TCGCCGGAGCCTCTATCAAGATGCCCCTGCGGAAGGCCTCAGCGAGCGTCAGCGGAAAGGATGG  1495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    85  TCGCCGGAGCCTCTATCAAGATTGCCCCTGCGGAAGGCCTCAGCGAGCGTCAGCGAAAGGATGG   144

Qy  1496  TCATCATCACCGGCCACCGGAGCCCACCCAGTTCAAGGGCCCAGGACCGGATCTTTGGGAAAAC  1555
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145  TCATCATCACCGGCCACCGGAGCCCACCCAGTTCAAGGGCCCAGGACCGGATCTTTGGGAAAAC   204

Qy  1556  TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCCATATCAGAG  1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205  TGAAAGAGGAAAACTTCTTTAACCCCAAAGAAGAAGTGGAAGCGCCATATCAGAG   264

Qy  1616  TGCCCCTTCCACAGCTGCCGCGGTGATTGGCAAAGGTGGCAAAGGTGGCAAGAGACCGCAAGAACTGC  1675
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265  TGCCCCTTCCACAGCTGCCGCGGTGATTGGCAAAGGTGGCAAAGGTGGCAAGAGACCGCAAGAACTGC   324

Qy  1676  AGAACTTAACCAGTGCAGAAGTCATCGTGCCCTGTGACCAGACGCCAGATGAAAATGAGG  1735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   325  AGAACTTAACCAGTGCAGAAGTCATCGTGCCTGTGACCAGACGCCAGATGAAAATGAGG   384

Qy  1736  AAGTGATGGTCAGAATTATCGGGACACTTCTTCTTGCTAGCCAGACTGCGACAGCGCAAGATCA  1795
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   385  AAGTGATGGTCAGAATTATCGGGACACTTCTTCTTGCTAGCCAGACTGCGACAGCGCAAGATCA   444

Qy  1796  GGGAAATTGTACAACAGTGAAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTTCAC  1855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445  GGGAAATTGTACAACAGTGAAGCAGAGCAGAAATACCCTCAGGGAGTCGCCTTCAC   504

Qy  1856  AGCGGCAGCAAGTGAGGCTCCACAGGCACCAGCAGCACCAGCACCGATGAATGTAGCCCTTCC  1915
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505  AGCGGCAGCAAGTGAGGCTCCACAGGCACCAGCAGCACCAGCACCGATGAATGTAGCCCTTCC   564

Qy  1916  AACACCTGACAGAATGAGACCAAAGCAGCCAGTCGGGAGCAAACCAAAGACCAT  1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565  AACACCTGACAGAATGAGACCAAAGACCAGCCAGTCGGGAGCAAACCAAAGACCAT   624

Qy  1976  CTGAGGAATGAGAAGTCTGCGAGGCCGGCGGCCTGCGAGGCCCTGAGAACCCCA  2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   625  CTGAGGAATGAGAAGTCTGCGAGGCCGGCGGCCTGCGAGGCCCTGAGAACCCCA   684
```

```
Qy  2343  TAGAAATAATATATCAACTAACTCCAATTTTTAATCAATTATTAATTAATTTTTT  2402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2580  TAGAAATAATATATCAAATAACTCAACTCCAATTTTTAATCAATTATTAATTTTTT  2639

Qy  2403  TTTCTTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT  2462
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2640  TTTCTTTTTAAAGAGAAAGCAGGCTTTCTAGACTTTAAAGAATAAAGTCTTTGGGAGGT  2699

Qy  2463  CTCACGGTGTGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCGAGGGAAATCT  2522
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2700  CTCACGGTGTGAGAGGAGGAGCTTTGAGGCCACCCGCACAAAAATTCACCCGAGGGAAATCT  2759

Qy  2523  CGTCGGAAGGACACTCACGGCAGTTCGGATCACCTGTGTATGTCAACAGAAGGGATACC  2582
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2760  CGTCGGAAGGACACTCACGGCAGTTCGGATCACCTGTGTATGTCAACAGAAGGGATACC  2819

Qy  2583  GTCTCCTTGAAGAGGAAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCACCCATTTC  2642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2820  GTCTCCTTGAAGAGGAAAACTCTGTCACTCCTCATGCCTGTCTAGCTCATCACCCATTTC  2879

Qy  2643  TCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATCTGCTATATAATTCTCTGTCTCT  2702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2880  TCTTTGCTTCACAGGTTTTAAACTGGTTTTTTGCATATCTGCTATATAATTCTCTGTCTCT  2939

Qy  2703  CTCTGTTTATCTCCCCTCCCCTCCCCTTCTTCCATCTCCAATTCTTTGAATT  2762
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2940  CTCTGTTTATCTCCCCTCCCCTCCCCTTCTTCCATCTCCAATTCTTTGAATT  2999

Qy  2763  TCCTCATCCCTCCAATCTCATTCTACGCACCCCCCCCCCCCCCAGGCAAAGCAGT  2822
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  3000  TCCTCATCCCTCCAATCTCATTCTACGCA-CCCCCCCCCCAGGCAAAGCAGT  3058

Qy  2823  GCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAAACACACAGCCTCAACT  2882
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3059  GCTCTGAGTATCACATCACACAAAAGGAACAAAAGCGAAAACACACAGCCTCAACT  3118

Qy  2883  TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG  2942
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3119  TACACTTGGTTACTCAAAGAACAAGAGTCAATGGTACTTGTCCTAGCGTTTTGGAAGAG  3178

Qy  2943  GAAAACAGGAACCCACCAAACACCAACCAACCAAACAAAGAAAAATTCCACAATGAA  3002
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3179  GAAAACAGGAACCCACCAAACACCAACCAACCAAACAAAGAAAAATTCCACAATGAA  3238

Qy  3003  AGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT  3062
          ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  3239  AGAATGTATTTTGTCTTTTGCATTTGGTGTATAAGCCATCAATATTCAGCAAAATGAT  3298

Qy  3063  TCCTTTCTTT-AAAAAAAAAAAATGTGGAGAAAGTAGAAATTACCAAGGTTGTTGGCCC  3121
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3299  TCCTTTCTTTAAAAAAAAAAAATGTGGAGAAAGTAGAAATTACCAAGGTTGTTGGCCC  3358

Qy  3122  AGGGCGTTAAATTCACAGATTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAG  3181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3359  AGGGCGTTAAATTCACAGATTTTTAACGAGAAAAACACACAGAAGAGCTACCTCAG  3418

Qy  3182  TGTTTTTACCTGACCTTGCTCTTGTCTCCTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGT  3241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3419  TGTTTTTACCTGACCTTGCTCTTGTGTCCTGTTTCCCTTAGAGATTTTGTAAAGCTGATAGT  3478

Qy  3242  TGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAATAAGATATCAACTG  3301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3479  TGGAGCATTTTTTATTTTTTAATAAAAATGAGTTGGAAAAATAAGATATCAACTG  3538

Qy  3302  CCAGCCTGGAGAAGTGACAGTCCAAGTGTGCAACAGCTGTTGTCTTCCGCTA  3361
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```